

OXFORD IB DIPLOMA PROGRAMME



VISIT...

ANSWERS

LANZAROTE
Caliente.COM

MATHEMATICS HIGHER LEVEL:
STATISTICS

COURSE COMPANION

Josip Harcet
Lorraine Heinrichs
Palmira Mariz Seiler
Marlene Torres Skoumal

OXFORD

1

Exploring further probability distributions

Skills check

- 1 a Mode (X) = -1, 1 because those two values share the highest probability of 0.3.

Median, $m = 1$ since $P(X \leq 0) = 0.4$ and $P(X \leq 1) = 0.7$

$$\mu = E(X) = \sum_{i=1}^6 x_i p_i = -1 \times 0.3 + 0 \times 0.1 + 1 \times 0.3 + 2 \times 0.1 + 3 \times 0.05 + 4 \times 0.15 = 0.95$$

$$\sigma = \sqrt{\text{Var}(X)} = \sqrt{\sum_{i=1}^6 x_i^2 p_i - \mu^2}$$

$$\begin{aligned} \sum_{i=1}^6 x_i^2 p_i &= (-1)^2 \times 0.3 + 0^2 \times 0.1 + 1^2 \times 0.3 \\ &\quad + 2^2 \times 0.1 + 3^2 \times 0.05 + 4^2 \times 0.15 \\ &= 3.85 \end{aligned}$$

$$\sigma = \sqrt{3.85 - 0.9025} = 1.72$$

Skills check			
	A	B	C
8			MinX
9			Q1X
10			MedianX...
11			Q3X
12			MaxX
D10			=1.

Skills check			
	A	B	C
2	0	0.1	$\bar{x}$
3	1	0.3	Σx
4	2	0.1	Σx^2
5	3	0.05	$s_x := s_{n-1}$
6	4	0.15	$s_x := s_{n-1}$
D6			=1.7168284713389

b

x_i	1	2	3	4
p_i	0.4	0.3	0.2	0.1

Mode (X) = 1 because it has the highest probability (0.4)

Median, $m = 2$ since $P(X \leq 1) = 0.4$ and $P(X \leq 2) = 0.7$

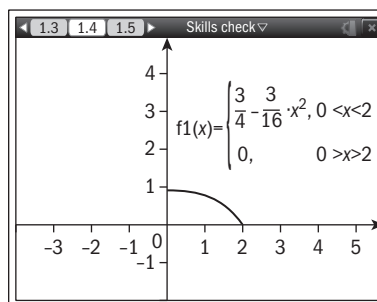
$$\begin{aligned} \mu = E(X) &= \sum_{i=1}^4 x_i p_i = 1 \times 0.4 + 2 \times 0.3 \\ &\quad + 3 \times 0.2 + 4 \times 0.1 = 2 \end{aligned}$$

$$\begin{aligned} \sigma &= \sqrt{\text{Var}(X)} = \sqrt{\sum_{i=1}^4 x_i^2 p_i - \mu^2} \\ &= \sqrt{1^2 \times 0.4 + 2^2 \times 0.3 + 3^2 \times 0.2 + 4^2 \times 0.1 - 2^2} \\ &= \sqrt{5 - 4} = 1 \end{aligned}$$

Skills check			
	A	B	C
8			MinX
9			Q1X
10			MedianX...
11			Q3X
12			MaxX
D10			=2.

Skills check			
	A	B	C
2	2	0.3	$\bar{x}$
3	3	0.2	Σx
4	4	0.1	Σx^2
5			$s_x := s_{n-1}$
6			$s_x := s_{n-1}$
D6			=1.

2 a



Mode (X) = 0, the part of parabola is opened downwards and it has its vertex at $\left(0, \frac{3}{4}\right)$.

$$\text{Median, } \int_0^m \left(\frac{3}{4} - \frac{3}{16}x^2\right) dx = \frac{1}{2} \Rightarrow$$

$$\frac{3}{4}m - \frac{1}{16}m^3 = \frac{1}{2} \Rightarrow m = 0.695$$

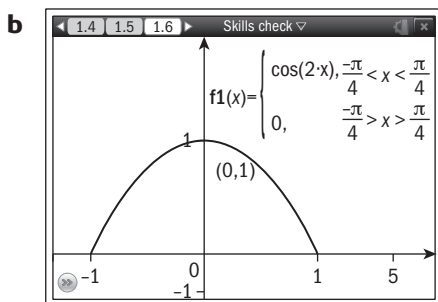
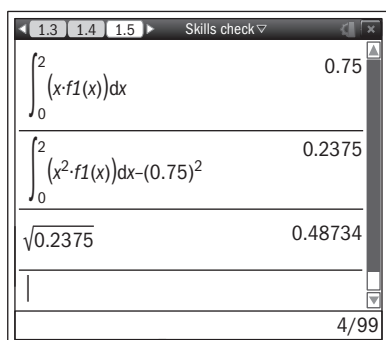
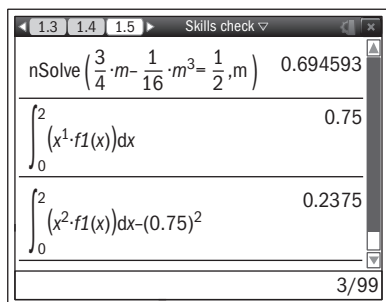
$$\mu = E(X) = \int_0^2 \left(\frac{3}{4}x - \frac{3}{16}x^3\right) dx$$

$$= \left[\frac{3}{8}x^2 - \frac{3}{64}x^4\right]_0^2 = \frac{3}{2} - \frac{3}{4} = \frac{3}{4}$$

$$\sigma = \sqrt{\text{Var}(X)} = \sqrt{\int_0^2 \left(\frac{3}{4}x^2 - \frac{3}{16}x^4\right) dx - \left(\frac{3}{4}\right)^2}$$

$$= \sqrt{\left[\frac{x^3}{4} - \frac{3x^5}{80} \right]_0^2} - \frac{9}{16} = \sqrt{2 - \frac{6}{5} - \frac{9}{16}}$$

$$= \sqrt{\frac{19}{80}} = 0.487$$

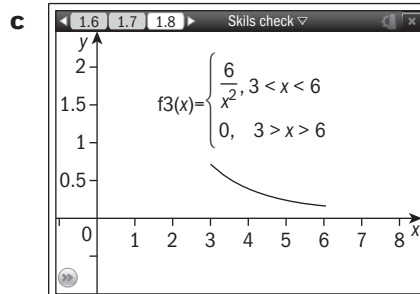
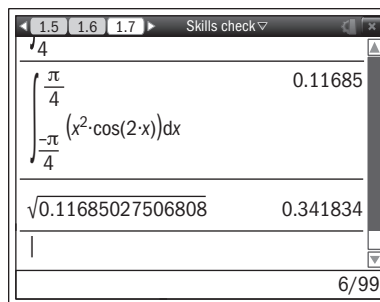
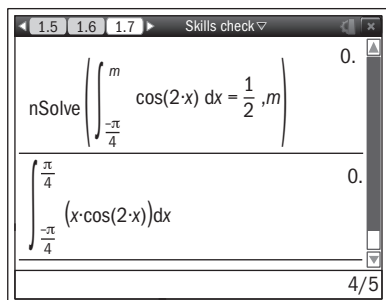


Mode (X) = 0, it has the maximum point at $(0, 1)$.

Median, $\int_{-\frac{\pi}{4}}^m f(x) dx = \frac{1}{2} \Rightarrow m = 0$

$$\mu = E(X) = \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} x f(x) dx = 0$$

$$\sigma = \sqrt{\text{Var}(X)} = \sqrt{\int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} x^2 f(x) dx - \mu^2} = 0.342$$

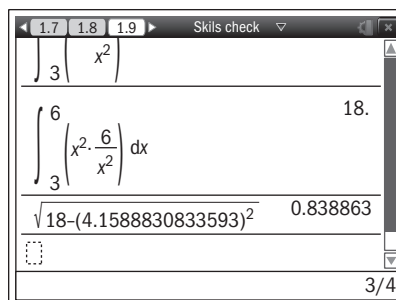
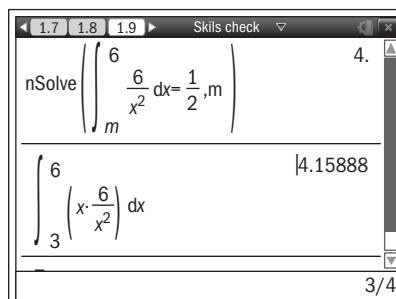


Mode (X) = 3 because it is a decreasing function on the interval $[3, 6]$.

Median, $\int_0^m f(x) dx = \frac{1}{2} \Rightarrow m = 4$

$$\mu = E(X) = \int_3^6 x f(x) dx = 4.16$$

$$\sigma = \sqrt{\text{Var}(X)} = \sqrt{\int_3^6 x^2 f(x) dx - \mu^2} = 0.839$$



3 a $1 - 0.5 + 0.25 - 0.125 + \dots \Rightarrow u_1 = 1, r = -0.5$

$$\text{Sum} = \frac{1}{1 - (-0.5)} = \frac{1}{1.5} = \frac{2}{3}$$

b $\sqrt{2} + 1 + \frac{\sqrt{2}}{2} + \frac{1}{2} \dots \Rightarrow u_1 = \sqrt{2}, r = \frac{1}{\sqrt{2}}$

$$\text{Sum} = \frac{\sqrt{2}}{1 - \frac{1}{\sqrt{2}}} = \frac{2}{\sqrt{2} - 1} \times \frac{\sqrt{2} + 1}{\sqrt{2} + 1} = 2 + 2\sqrt{2}$$

$$\begin{aligned}
 \text{4 a } f(x) &= \frac{1}{2-x}, x \neq 2 \\
 \Rightarrow f'(x) &= \frac{-1 \times -1}{(2-x)^2} = \frac{1}{(2-x)^2}, x \neq 2 \\
 \Rightarrow \int f(x) dx &= \int \frac{dx}{2-x} = -\ln(2-x) + c, x < 2
 \end{aligned}$$

$$\begin{aligned}
 \text{b } f(x) &= e^{3x+1} \Rightarrow f'(x) = 3e^{3x+1} \\
 \Rightarrow \int f(x) dx &= \int e^{3x+1} dx = \frac{1}{3} e^{3x+1} + c
 \end{aligned}$$

$$\begin{aligned}
 \text{c } f(x) &= \frac{3}{2} \sin \frac{\pi-2x}{3} \\
 \Rightarrow f'(x) &= \frac{3}{2} \cos \frac{\pi-2x}{3} \times \left(-\frac{2}{3}\right) = -\cos \frac{\pi-2x}{3} \\
 \Rightarrow \int f(x) dx &= \int \frac{3}{2} \sin \frac{\pi-2x}{3} dx \\
 &= \frac{3}{2} \times -\frac{1}{\frac{2}{3}} \cos \frac{\pi-2x}{3} + c = \frac{9}{4} \cos \frac{\pi-2x}{3} + c
 \end{aligned}$$

$$\begin{aligned}
 \text{d } f(x) &= (x^2 - 2)^2 \\
 \Rightarrow f'(x) &= 2(x^2 - 2) \times 2x = 4x(x^2 - 2) \\
 \Rightarrow \int f(x) dx &= \int (x^2 - 2)^2 dx \\
 &= \int (x^4 - 4x^2 + 4) dx = \frac{x^5}{5} - \frac{4x^3}{3} + 4x + c
 \end{aligned}$$

Exercise 1A

$$\text{1 a } \frac{k}{12} + \frac{1+k}{12} + \frac{2+k}{12} = 1 \Rightarrow 3 + 3k = 12 \Rightarrow k = 3$$

$$\text{b }$$

$X = x$	0	1	2
$F(x)$	$\frac{3}{12}$	$\frac{7}{12}$	$\frac{12}{12}$

$$\Rightarrow F(x) = \begin{cases} 0, & x < 0 \\ \frac{x^2 + 7x + 6}{24}, & x = 0, 1, 2 \\ 1, & x > 2 \end{cases}$$

$$\begin{aligned}
 \text{2 a } \frac{a-2}{20} + \frac{a-4}{20} + \frac{a-6}{20} + \frac{a-8}{20} &= 1 \\
 \Rightarrow 4a - 20 &= 20 \Rightarrow a = 10
 \end{aligned}$$

$$\text{b }$$

$X = x$	1	2	3	4
$F(x)$	$\frac{8}{20}$	$\frac{14}{20}$	$\frac{18}{20}$	$\frac{20}{20}$

$$\Rightarrow F(x) = \begin{cases} 0, & x < 1 \\ \frac{9x - x^2}{20}, & x = 1, 2, 3, 4 \\ 1, & x > 4 \end{cases}$$

$$\text{c } P(X \leq 2) = F(2) = \frac{14}{20} = \frac{7}{10}$$

$$\begin{aligned}
 \text{3 a } P(X = 0) &= F(0) = \frac{1}{6}, \\
 P(X = 1) &= F(1) - F(0) = \frac{1}{2} - \frac{1}{6} = \frac{2}{6}, \\
 P(X = 2) &= F(2) - F(1) = 1 - \frac{1}{2} = \frac{1}{2} \\
 &= \frac{3}{6} \Rightarrow P(X = x) = \begin{cases} \frac{x+1}{6}, & x = 0, 1, 2 \\ 0, & \text{otherwise} \end{cases}
 \end{aligned}$$

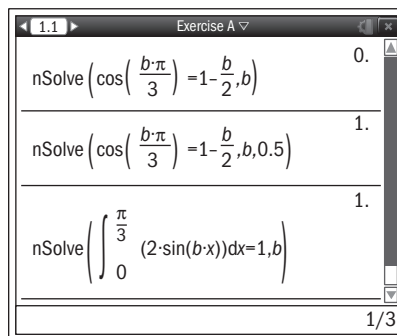
b The modal value of the variable X is 2 since it has the highest probability, $P(X = 2) = \frac{1}{2}$.

$$\begin{aligned}
 \text{4 a } P(X = 1) &= F(1) = \frac{1}{25} \\
 P(X = 3) &= F(3) - F(1) = \frac{3}{25} \\
 P(X = 5) &= F(5) - F(3) = \frac{5}{25} \\
 P(X = 7) &= F(7) - F(5) = \frac{7}{25} \\
 P(X = 9) &= F(9) - F(7) = \frac{9}{25} \\
 P(X = x) &= \begin{cases} \frac{x}{25}, & x = 1, 3, 5, 7, 9 \\ 0, & \text{otherwise} \end{cases}
 \end{aligned}$$

b Median, $m = 7$ since

$$F(5) = \frac{9}{25} < \frac{1}{2} \text{ and } F(7) = \frac{16}{25} > \frac{1}{2}$$

$$\begin{aligned}
 \text{5 a } \int_0^{\frac{\pi}{3}} 2 \sin bx \, dx &= 1 \Rightarrow 2 \left[-\frac{1}{b} \cos bx \right]_0^{\frac{\pi}{3}} = 1 \\
 \Rightarrow \cos\left(\frac{b\pi}{3}\right) - \cos 0 &= -\frac{b}{2} \\
 \Rightarrow \cos\left(\frac{b\pi}{3}\right) &= 1 - \frac{b}{2} \Rightarrow b = 1
 \end{aligned}$$



$$\begin{aligned}
 \text{b } F(x) &= \int_{-\infty}^x f(t) dt \Rightarrow F(x) = \int_0^x 2 \sin t \, dt \\
 &= [-2 \cos t]_0^x = -2 \cos x + 2 \\
 \Rightarrow F(x) &= \begin{cases} 0, & x < 0 \\ 2 - 2 \cos x, & 0 \leq x \leq \frac{\pi}{3} \\ 1, & x > \frac{\pi}{3} \end{cases}
 \end{aligned}$$

$$\begin{aligned} \text{c } P\left(X \geq \frac{\pi}{6}\right) &= 1 - P\left(X \leq \frac{\pi}{6}\right) = 1 - F\left(\frac{\pi}{6}\right) \\ &= 1 - 2 + 2\cos\left(\frac{\pi}{6}\right) = \sqrt{3} - 1 = 0.732 \end{aligned}$$

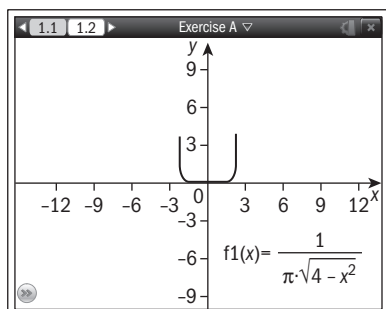
- 6 a Since the function is not defined for the values of -2 and 2 we need to use limits.

$$\begin{aligned} \lim_{b \rightarrow 2} \left(\int_{-b}^b \frac{dx}{\pi \sqrt{4-x^2}} \right) &= \lim_{b \rightarrow 2} \left[\frac{1}{\pi} \arcsin \frac{x}{2} \right]_{-b}^b \\ &= \frac{1}{\pi} \left[\arcsin \frac{x}{2} \right]_{-2}^2 = \frac{1}{\pi} (\arcsin 1 - \arcsin(-1)) \\ &= \frac{1}{\pi} \left(\frac{\pi}{2} + \frac{\pi}{2} \right) = 1 \end{aligned}$$

The function is well defined if $\int_{-\infty}^{+\infty} f(x) dx = 1$

- b The modal value is the value of the random variable X where probability reaches its maximum.

$$\begin{aligned} f'(x) &= \frac{1}{\pi} \left(-\frac{1}{2} \right) (4-x^2)^{-\frac{3}{2}} (-2x) \\ &= \frac{x}{\pi(4-x^2)^{\frac{3}{2}}}, \quad f'(x) = 0 \Rightarrow x = 0 \end{aligned}$$



By looking at the graph of the probability function we notice that the graph has only one extreme point and that is the minimum point at 0, therefore the modal value doesn't exist.

Exercise 1B

- 1 a $X \sim \text{Geo}(0.6) \Rightarrow P(X=2) = 0.4 \times 0.6 = 0.24$
 b $X \sim \text{Geo}(0.14) \Rightarrow P(X=3) = 0.86^2 \times 0.14 = 0.104$
 c $X \sim \text{Geo}(0.5) \Rightarrow P(X=4) = 0.5^3 \times 0.5 = 0.0625$
 d $X \sim \text{Geo}(0.88) \Rightarrow P(X=5) = 0.12^4 \times 0.88 = 0.000182$

geomPdf(0.6,2)	0.24
geomPdf(0.14,3)	0.103544
geomPdf(0.5,4)	0.0625
geomPdf(0.88,5)	0.000182

- 2 a $X \sim \text{Geo}(0.25) \Rightarrow P(X \leq 4) = 0.684$
 b $X \sim \text{Geo}(0.7) \Rightarrow P(X > 6) = 1 - P(X \leq 6) = 0.000729$
 c $X \sim \text{Geo}(0.3) \Rightarrow P(5 \leq X \leq 7) = P(X \leq 7) - P(X \leq 4) = 0.158$
 d $X \sim \text{Geo}(0.991) \Rightarrow P(1 < X \leq 7) = P(X \leq 7) - P(X \leq 1) = 0.009$

geomCdf(0.25,1,4)	0.683594
1-geomCdf(0.7,1,6)	0.000729
geomCdf(0.3,1,7)-geomCdf(0.3,1,4)	0.157746
geomCdf(0.991,1,7)-geomCdf(0.991,1,1)	0.009

- 3 $X \sim \text{Geo}(0.73) \Rightarrow P(X \leq 3) = 0.980$
 4 $X \sim \text{Geo}(0.92)$
 a $P(X=5) = 0.0000377$
 b $P(X \geq 4) = 1 - P(X \leq 3) = 0.000512$
 5 $X \sim \text{Geo}(0.15)$

geomCdf(0.73,1,3)	0.980317
geomPdf(0.92,5)	0.000038
1-geomCdf(0.92,1,3)	0.000512
geomPdf(0.15,4)	0.092119
1-geomCdf(0.15,1,6)	0.37715

- a $P(X=4) = 0.0921$
 b $P(X \geq 7) = 1 - P(X \leq 6) = 0.377$
 6 $X \sim \text{Geo}(p) \Rightarrow P(X > k) = 1 - P(X \leq k)$

$$= 1 - \sum_{n=k+1}^{\infty} q^{n-1} p = 1 - p \underbrace{\sum_{n=1}^k q^{n-1}}_{\text{inf. geom. seq.}}$$

$$= 1 - p \frac{1-q^k}{1-q} = 1 - p \frac{1-q^k}{p} = 1 - (1-q^k) = q^k$$

OR

$$\begin{aligned} X \sim \text{Geo}(p) \Rightarrow P(X > k) &= \sum_{n=k+1}^{\infty} q^{n-1} p = q^k p \times \underbrace{\sum_{n=1}^{\infty} q^{n-1}}_{\text{inf. geom. seq.}} = q^k p \frac{1}{1-q} = q^k p \frac{1}{p} = q^k \end{aligned}$$

Investigation 1

$$X \sim \text{Geo}(p)$$

a $p = 0.4 \Rightarrow P(X > 5 | X > 3)$

$$= \frac{P((X > 5) \cap (X > 3))}{P(X > 3)} = \frac{1 - P(X \leq 5)}{1 - P(X \leq 3)} = 0.36$$

$$P(X > 2) = 1 - P(X \leq 2) = 0.36$$

b $p = 0.7 \Rightarrow P(X > 6 | X > 2)$

$$= \frac{P((X > 6) \cap (X > 2))}{P(X > 2)} = \frac{1 - P(X \leq 6)}{1 - P(X \leq 2)} = 0.0081$$

$$P(X > 4) = 1 - P(X \leq 4) = 0.0081$$

c $p = 0.12 \Rightarrow P(X > 12 | X > 5)$

$$= \frac{P((X > 12) \cap (X > 5))}{P(X > 5)} = \frac{1 - P(X \leq 12)}{1 - P(X \leq 5)} = 0.409$$

$$P(X > 7) = 1 - P(X \leq 7) = 0.409$$

Expression	Result
1-geomCdf(0.4,5)	0.36
1-geomCdf(0.4,3)	
1-geomCdf(0.4,2)	0.36
1-geomCdf(0.7,6)	0.0081
1-geomCdf(0.7,2)	
1-geomCdf(0.7,4)	0.0081

Expression	Result
1-geomCdf(0.12,12)	0.408676
1-geomCdf(0.12,5)	
1-geomCdf(0.12,7)	0.408676

$$X \sim \text{Geo}(p); m, n \in \mathbb{Z}^+; m < n$$

$$\Rightarrow P(X > n | X > m) = \frac{P((X > n) \cap (X > m))}{P(X > m)}$$

$$= \frac{P(X > n)}{P(X > m)} = \frac{q^n}{q^m} = q^{n-m} = P(X > n - m)$$

Exercise 1C

1 For 1B, Question 1:

a $X \sim \text{Geo}(0.6) \Rightarrow E(X) = \frac{1}{0.6} = 1.67,$

$$\text{Var}(X) = \frac{0.4}{0.6^2} = 1.11$$

b $X \sim \text{Geo}(0.14) \Rightarrow E(X) = \frac{1}{0.14} = 7.14,$

$$\text{Var}(X) = \frac{0.86}{0.14^2} = 43.9$$

c $X \sim \text{Geo}(0.5) \Rightarrow E(X) = \frac{1}{0.5} = 2,$

$$\text{Var}(X) = \frac{0.5}{0.5^2} = 2$$

d $X \sim \text{Geo}(0.88) \Rightarrow E(X) = \frac{1}{0.88} = 1.14,$

$$\text{Var}(X) = \frac{0.12}{0.88^2} = 0.155$$

For 1B, Question 2:

a $X \sim \text{Geo}(0.25) \Rightarrow E(X) = \frac{1}{0.25} = 4,$

$$\text{Var}(X) = \frac{0.75}{0.25^2} = 12$$

b $X \sim \text{Geo}(0.7) \Rightarrow E(X) = \frac{1}{0.7} = 1.43,$

$$\text{Var}(X) = \frac{0.3}{0.7^2} = 0.612$$

c $X \sim \text{Geo}(0.3) \Rightarrow E(X) = \frac{1}{0.3} = 3.33,$

$$\text{Var}(X) = \frac{0.7}{0.3^2} = 7.78$$

d $X \sim \text{Geo}(0.991) \Rightarrow E(X) = \frac{1}{0.991} = 1.01,$

$$\text{Var}(X) = \frac{0.009}{0.991^2} = 0.00916$$

2 $X \sim \text{Geo}(0.73)$

a $E(X) = \frac{1}{0.73} = 1.37$

b $\sigma = \sqrt{\text{Var}(X)} = \sqrt{\frac{0.27}{0.73^2}} = 0.712$

$$x_{\max} = E(X) + 3\sigma = \frac{1}{0.73} + 3 \times \sqrt{\frac{0.27}{0.73^2}} = 3.51$$

Mario must make four shots to destroy the balloon.

$\frac{1}{0.73}$	1.36986
$\sqrt{\frac{0.27}{(0.73)^2}}$	0.711802
$\frac{1}{0.73} + 3 \times \sqrt{\frac{0.27}{(0.73)^2}}$	3.50527

3 $X \sim \text{Geo}(0.54) \Rightarrow E(X) = \frac{1}{0.54} = 1.85$

$$\sigma = \sqrt{\text{Var}(X)} = \sqrt{\frac{0.46}{0.54^2}} = 1.26$$

$$x_{\max} = E(X) + 3\sigma = \frac{1}{0.54} + 3 \times \sqrt{\frac{0.46}{0.54^2}} = 5.62$$

Therefore, we must select 6 students at random to ensure that one of them will be familiar with the election procedure.

$\frac{1}{0.54}$	1.85185
$\frac{1}{0.54} + 3 \cdot \sqrt{\frac{0.46}{(0.54)^2}}$	5.61981

Exercise 1D

1 a $X \sim \text{NB}(1, 0.2) \Rightarrow P(X = 2)$

$$= \binom{2-1}{1-1} 0.8 \times 0.2 = 0.16$$

b $X \sim \text{NB}(3, 0.5) \Rightarrow P(X = 4)$

$$= \binom{4-1}{3-1} 0.5 \times 0.5^3 = 0.188$$

c $X \sim \text{NB}(7, 0.8) \Rightarrow P(X = 9)$

$$= \binom{9-1}{7-1} 0.2^2 \times 0.8^7 = 0.235$$

d $X \sim \text{NB}(23, 0.77) \Rightarrow P(X = 32)$

$$= \binom{32-1}{23-1} 0.23^9 \times 0.77^{32} = 0.00847$$

$nCr(1,0) \cdot 0.8 \cdot 0.2$	0.16
$nCr(3,2) \cdot 0.5 \cdot (0.5)^3$	0.1875
$nCr(8,6) \cdot (0.2)^2 \cdot (0.8)^7$	0.234881
$nCr(31,22) \cdot (0.23)^9 \cdot (0.77)^{32}$	0.008467

2 a $X \sim \text{NB}(0.25, 3) \Rightarrow P(X \leq 4)$

$$= P(X = 3) + P(X = 4) = \frac{13}{256} = 0.508$$

$\sum_{k=3}^4 (nCr(k-1, 3-1) \cdot (0.75)^{k-3} \cdot (0.25)^3)$	0.050781
--	----------

b $X \sim \text{NB}(0.5, 2) \Rightarrow P(X > 6)$

$$= 1 - \sum_{k=2}^6 P(X = k) = \frac{7}{64} = 0.109$$

$1 - \sum_{k=2}^6 (nCr(k-1, 2-1) \cdot (0.5)^{k-2} \cdot (0.5)^2)$	0.109375
--	----------

c $X \sim \text{NB}(0.82, 4) \Rightarrow P(5 \leq X \leq 7)$

$$= \sum_{k=5}^7 P(X = k) = 0.525$$

$\sum_{k=5}^7 (nCr(k-1, 4-1) \cdot (0.18)^{k-4} \cdot (0.82)^4)$	0.524751
--	----------

d $X \sim \text{NB}(0.43, 5) \Rightarrow P(8 < X \leq 11)$

$$= \sum_{k=9}^{11} P(X = k) = 0.326$$

$\sum_{k=9}^{11} (nCr(k-1, 5-1) \cdot (0.57)^{k-5} \cdot (0.43)^5)$	0.325959
---	----------

3 a $P(X = 3) = \frac{24}{125} \Rightarrow \binom{3-1}{2-1} (1-p) \times p^2$

$$= \frac{24}{125} \Rightarrow p^2 - p^3 = \frac{12}{125} \Rightarrow \begin{cases} p_1 = -0.274 \\ p_2 = \frac{2}{5} \\ p_3 = 0.874 \end{cases}$$

Thus $p = \frac{2}{5} = 0.4$

$$\begin{aligned} \text{b } X &\sim \text{NB}\left(2, \frac{2}{5}\right) \Rightarrow P(3 \leq X \leq 5) \\ &= \sum_{k=3}^5 P(X = k) = \frac{1572}{3125} = 0.503 \end{aligned}$$

$$4 \text{ a } X \sim \text{NB}\left(2, \frac{1}{4}\right)$$

$$\begin{aligned} \text{b } X &\sim \text{NB}\left(2, \frac{1}{4}\right) \Rightarrow P(X = x) = \frac{3}{32} \\ &\Rightarrow \binom{x-1}{2-1} \left(\frac{3}{4}\right)^{x-2} \times \left(\frac{1}{4}\right)^2 = \frac{3}{32} \\ &\Rightarrow (x-1) \frac{3^{x-2}}{4^x} = \frac{3}{32} \Rightarrow x = 3 \end{aligned}$$

$$\begin{aligned} \text{c } X &\sim \text{NB}\left(2, \frac{1}{4}\right) \Rightarrow P(X \leq 5) \\ &= \sum_{k=2}^5 P(X = k) = \frac{47}{128} \end{aligned}$$

$$5 \text{ a } X \sim \text{NB}(0.85, 4) \Rightarrow P(X = 5) = 0.313$$

$$\begin{aligned} \text{b } X &\sim \text{NB}(0.85, 4) \Rightarrow P(X \geq 7) \\ &= 1 - P(X \leq 6) = 0.0473 \end{aligned}$$

$$6 \text{ a } X \sim \text{NB}(0.92, 5) \Rightarrow P(X = 6) = 0.264$$

$$\text{b } X \sim \text{NB}(0.92, 5) \Rightarrow P(X \leq 12) = 0.999999 \approx 1$$

We can say that it is almost certain that he will not need to interview more than a dozen students.

$$7 \text{ a } X \sim \text{NB}(0.85, 3) \Rightarrow P(X = 5) = 0.0829$$

$$\begin{aligned} \text{b } X &\sim \text{NB}(0.85, 3) \Rightarrow P(X \geq 7) \\ &= 1 - P(X \leq 6) = 0.589 \end{aligned}$$

Exercise 1E

$$1 \quad P(X = x) = \underbrace{\binom{4}{x} \left(\frac{1}{2}\right)^x}_{\text{heads}} \underbrace{\left(\frac{1}{2}\right)^{4-x}}_{\text{tails}} = \binom{4}{x} \left(\frac{1}{2}\right)^4$$

x_i	0	1	2	3	4
p_i	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

$$G(t) = \frac{1}{16} + \frac{1}{4}t + \frac{3}{8}t^2 + \frac{1}{4}t^3 + \frac{1}{16}t^4$$

$$2$$

x_i	1	2	3	...	k	...
p_i	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{25}{216}$	...	$\underbrace{\left(\frac{5}{6}\right)^{k-1}}_{\text{not one}} \times \underbrace{\frac{1}{6}}_{\text{one}} = \frac{5^{k-1}}{6^k}$	...

$$\begin{aligned} G(t) &= \frac{1}{6}t + \frac{5}{36}t^2 + \frac{25}{216}t^3 + \dots \\ &+ \frac{5^{k-1}}{6^k}t^k + \dots = \frac{1}{6}t \sum_{k=0}^{\infty} \left(\frac{5}{6}\right)^k \\ &\quad \text{inf. geom. sequence} \\ &= \frac{1}{6}t \times \frac{1}{1 - \frac{5}{6}} = \frac{1}{6}t \times \frac{6}{6-5} = \frac{t}{6-5t} \\ \left|\frac{5}{6}t\right| < 1 &\Rightarrow -1 < \frac{5}{6}t < 1 \Rightarrow -\frac{6}{5} < t < \frac{6}{5} \end{aligned}$$

$$\begin{aligned} 3 \text{ a } G(t) &= \frac{2}{3-t} = \frac{\frac{2}{3}}{1 - \frac{1}{3}t} = \frac{2}{3} + \frac{2}{9}t + \frac{2}{27}t^2 \\ &+ \frac{2}{81}t^3 + \dots + \frac{2}{3^{k+1}}t^k + \dots \Rightarrow P(X = 0) = \frac{2}{3} \end{aligned}$$

$$\text{b } P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{2}{3} + \frac{2}{9} = \frac{8}{9}$$

$$\begin{aligned} \text{c } P(X \geq 3) &= 1 - P(X \leq 2) \\ &= 1 - \left(\frac{2}{3} + \frac{2}{9} + \frac{2}{27} \right) = 1 - \frac{26}{27} = \frac{1}{27} \end{aligned}$$

$$\begin{aligned} \text{d } P(X \geq k) &= 1 - P(X \leq k-1) \\ &= 1 - \left(\frac{2}{3} + \frac{2}{9} + \dots + \frac{2}{3^k} \right) \\ &= 1 - \frac{\frac{2}{3} \left(1 - \frac{1}{3^k} \right)}{1 - \frac{1}{3}} = 1 - 1 + \frac{1}{3^k} = \frac{1}{3^k} \end{aligned}$$

4 a $X \sim B(1, p)$

x_i	0	1
p_i	$1-p$	p

$$G(t) = 1 - p + pt$$

$$G'(t) = p, G''(t) = 0$$

$$E(X) = G'(1) = p,$$

$$\begin{aligned} \text{Var}(X) &= G''(1) + G'(1)(1 - G'(1)) \\ &= 0 + p(1 - p) = p(1 - p) [= pq] \end{aligned}$$

b $X \sim \text{NB}(r, p), G(t) = \left(\frac{pt}{1-qt} \right)^r$

$$\begin{aligned} G'(t) &= r \left(\frac{pt}{1-qt} \right)^{r-1} \frac{p(1-qt) - pt \times (-q)}{(1-qt)^2} \\ &= rp \left(\frac{pt}{1-qt} \right)^{r-1} \frac{1}{(1-qt)^2} = rp^r t^{r-1} (1-qt)^{-(r+1)} \end{aligned}$$

$$G'(1) = rp^r p^{-(r+1)} = \frac{r}{p}$$

$$\begin{aligned} G''(t) &= rp^r (r-1)t^{r-2} (1-qt)^{-(r+1)} \\ &\quad + rp^r t^{r-1} \times (-(r+1))(1-qt)^{-(r+2)} \times (-q) \\ &= rp^r t^{r-2} (1-qt)^{-(r+2)} ((r-1)(1-qt) + (r+1)qt) \end{aligned}$$

$$\begin{aligned} G''(1) &= rp^r \times p^{-r-2} ((r-1)p + (r+1)q) \\ &= \frac{r}{p^2} \left(-p + \overbrace{rp + rq}^r + q \right) = \frac{r}{p^2} (-p + r + q) \end{aligned}$$

$$E(X) = G'(1) = \frac{r}{p},$$

$$\begin{aligned} \text{Var}(X) &= G''(1) + G'(1)(1 - G'(1)) \\ &= \frac{r}{p^2} (-p + r + q) + \frac{r}{p} \left(1 - \frac{r}{p} \right) \\ &= \frac{-\cancel{rp} + \cancel{r^2} + \cancel{rq} + rp - \cancel{r^2}}{p^2} = \frac{rq}{p^2} \end{aligned}$$

5 For question 1:

$$G(t) = \frac{1}{16} + \frac{1}{4}t + \frac{3}{8}t^2 + \frac{1}{4}t^3 + \frac{1}{16}t^4,$$

$$G'(t) = \frac{1}{4} + \frac{3}{4}t + \frac{3}{4}t^2 + \frac{1}{4}t^3 \Rightarrow G'(1) = 2$$

$$G''(t) = \frac{3}{4} + \frac{3}{2}t + \frac{3}{4}t^2 \Rightarrow G''(1) = 3$$

$$\begin{aligned} E(X) &= G'(1) = 2; \text{Var}(X) = G''(1) \\ &\quad + G'(1)(1 - G'(1)) = 3 + 2(1 - 2) = 3 - 2 = 1 \end{aligned}$$

CH1 Exercise E	
$f1(1)$	1
$\frac{d}{dx}(f1(x)) _{x=1}$	2
$\frac{d^2}{dx^2}(f1(x)) _{x=1}$	3
	3/99

For question 2:

$$\begin{aligned} G(t) &= \frac{t}{6-5t}, G'(t) = \frac{6-5t-t \times (-5)}{(6-5t)^2} \\ &= 6(6-5t)^{-2} \Rightarrow G'(1) = 6 \end{aligned}$$

$$\begin{aligned} G''(t) &= -12(6-5t)^{-3} \times (-5) \\ &= 60(6-5t)^{-2} \Rightarrow G''(1) = 60 \end{aligned}$$

$$\begin{aligned} E(X) &= G'(1) = 6; \text{Var}(X) = G''(1) \\ &\quad + G'(1)(1 - G'(1)) = 60 + 6(1 - 6) = 60 - 30 = 30 \end{aligned}$$

CH1 Exercise E	
$f2(1)$	1
$\frac{d}{dx}(f2(x)) _{x=1}$	6
$\frac{d^2}{dx^2}(f2(x)) _{x=1}$	60
	6/99

For question 3:

$$\begin{aligned} G(t) &= \frac{2}{3-t}, G'(t) = 2(3-t)^{-2} \\ &\Rightarrow G'(1) = 2(3-1)^{-2} = \frac{1}{2} \end{aligned}$$

$$G''(t) = 4(3-t)^{-3} \Rightarrow G''(1) = 4(3-1)^{-3} = \frac{1}{2}$$

$$E(X) = G'(1) = \frac{1}{2},$$

$$\begin{aligned} \text{Var}(X) &= G''(1) + G'(1)(1 - G'(1)) \\ &= \frac{1}{2} + \frac{1}{2} \left(1 - \frac{1}{2} \right) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4} \end{aligned}$$

$\frac{d}{dx}(f_3(x)) _{x=1}$	1
$\frac{d^2}{dx^2}(f_3(x)) _{x=1}$	$\frac{1}{2}$
$\frac{d}{dx}(f_3(x)) _{x=1}$	$\frac{1}{2}$
$\frac{d^2}{dx^2}(f_3(x)) _{x=1}$	$\frac{1}{2}$
	9/99

6 a $P(X=1) = \frac{4}{7}; P(X=4) = \frac{3}{7} \times \frac{1}{3} \times \frac{3}{7} \times \frac{2}{3} = \frac{2}{49}$

b $G(t) = \frac{4}{7}t + \frac{3}{7} \times \frac{2}{3}t^2 + \frac{3}{7} \times \frac{1}{3} \times \frac{4}{7}t^3$
 $+ \frac{3}{7} \times \frac{1}{3} \times \frac{3}{7} \times \frac{2}{3}t^4 + \frac{3}{7} \times \frac{1}{3} \times \frac{3}{7} \times \frac{1}{3} \times \frac{4}{7}t^5$
 $+ \frac{3}{7} \times \frac{1}{3} \times \frac{3}{7} \times \frac{1}{3} \times \frac{2}{3}t^6 + \dots$
 $= \frac{4}{7}t \left(1 + \frac{1}{7}t^2 + \frac{1}{49}t^4 + \dots\right)$
 $+ \frac{2}{7}t^2 \left(1 + \frac{1}{7}t^2 + \frac{1}{49}t^4 + \dots\right)$
 $= \left(\frac{4}{7}t + \frac{2}{7}t^2\right) \left(1 + \frac{1}{7}t^2 + \frac{1}{49}t^4 + \dots\right)$
 $= \frac{4t + 2t^2}{7} \times \frac{1}{1 - \frac{1}{7}t^2} = \frac{4t + 2t^2}{7} \times \frac{7}{7 - t^2} = \frac{4t + 2t^2}{7 - t^2}$

Verifying: $G(1) = \frac{4 + 2}{7 - 1} = \frac{6}{6} = 1$

c $G(t) = \frac{4t + 2t^2}{7 - t^2}, G'(t) = \frac{(4 + 4t)(7 - t^2) + 2t(4t + 2t^2)}{(7 - t^2)^2}$
 $= \frac{4(7 + 7 - t^2 - t^2) + 2t^2 + t^3}{(7 - t^2)^2} = \frac{4(7 + 7t + t^2)}{(7 - t^2)^2}$

$E(X) = G'(1) = \frac{4(7 + 7 + 1)}{(7 - 1)^2} = \frac{5}{3}$

The expected number of shots before the game is over is 2.

d $G'(t) = 4(7 + 7t + t^2)(7 - t^2)^{-2}$
 $G''(t) = 4(7 + 2t)(7 - t^2)^{-2}$
 $- 2(7 - t^2)^{-3} \times (-2t) \times 4(7 + 7t + t^2)$
 $= \frac{4((7 + 2t)(7 - t^2) + 4t(7 + 7t + t^2))}{(7 - t^2)^3}$
 $= \frac{4(49 + 14t - 7t^2 - 2t^3 + 28t + 28t^2 + 4t^3)}{(7 - t^2)^3}$
 $= \frac{4(49 + 42t + 21t^2 + 2t^3)}{(7 - t^2)^3}$
 $\Rightarrow G''(1) = \frac{4(49 + 42 + 21 + 2)}{(7 - 1)^3} = \frac{19}{9}$

$\frac{d}{dx}(f_4(x)) _{x=1}$	1
$\frac{d}{dx}(f_4(x)) _{x=1}$	$\frac{5}{3}$
$\frac{d^2}{dx^2}(f_4(x)) _{x=1}$	$\frac{19}{9}$
	12/99

$\text{Var}(X) = G''(1) + G'(1)(1 - G'(1))$
 $= \frac{19}{9} + \frac{5}{3} \left(1 - \frac{5}{3}\right) = \frac{19}{9} - \frac{10}{9} = 1$

$x_{\max} = E(X) + 3\sigma = \frac{5}{3} + 3 \times \sqrt{1} = \frac{14}{3} = 4.67$

Therefore, the maximum number of shots to be made before the game is over is 5.

Exercise 1F

1 a $G_{X+Y}(t) = G_X(t) \times G_Y(t) = \left(\frac{1+3t}{4}\right)^2$
 $\times \left(\frac{2+t}{3}\right)^2 = \left(\frac{2+7t+3t^2}{12}\right)^2$

b $G_{X+Y}(t) = \left(\frac{2+7t+3t^2}{12}\right)^2 = \frac{4}{144} + \frac{28}{144}t + \dots$
 $\Rightarrow P(X+Y \leq 1) = \frac{4}{144} + \frac{28}{144} = \frac{32}{144} = \frac{2}{9}$

c $G'_{X+Y}(t) = \frac{1}{144} \times 2(2+7t+3t^2)(7+6t)$
 $\Rightarrow E(X+Y) = G'_{X+Y}(1) = \frac{1}{72} \times 12 \times 13 = \frac{13}{6}$

2 a $G_{X+Y}(t) = G_X(t) \times G_Y(t) = \left(\frac{t}{2-t}\right)^3 \times \left(\frac{t}{3-2t}\right)^3$
 $= \left(\frac{t^2}{6-7t+2t^2}\right)^3$

b $E(X+Y) = G'(1) = 15$

c $\text{Var}(X+Y) = G''(1) + G'(1)(1 - G'(1))$
 $= 234 + 15 \times (-14) = 234 - 210 = 24$

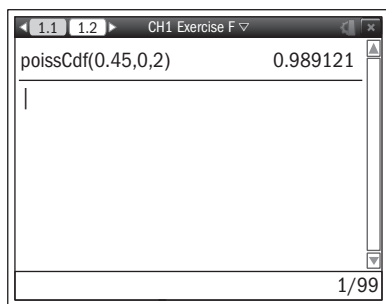
$\frac{d}{dt} \left(\left(\frac{t}{2-t} \right)^3 \cdot \left(\frac{t}{3-2t} \right)^3 \right) _{t=1}$	15
$\frac{d^2}{dt^2} \left(\left(\frac{t}{2-t} \right)^3 \cdot \left(\frac{t}{3-2t} \right)^3 \right) _{t=1}$	234
	2/99

3 a $G_X(t) = e^{0.25(t-1)}, G_Y(t) = e^{0.15(t-1)}, G_Z(t) = e^{0.05(t-1)}$

b $G_{X+Y+Z}(t) = G_X(t) \times G_Y(t) \times G_Z(t)$
 $= e^{0.25(t-1)} \times e^{0.15(t-1)} \times e^{0.05(t-1)} = e^{0.45(t-1)}$

$G_{X+Y+Z}(t) = e^{0.45(t-1)} \Rightarrow X+Y+Z \sim \text{Po}(0.45)$

$P(X+Y+Z \leq 2) = 0.989$



$$4 \quad X \sim \text{Geo}(p), G_X(t) = \frac{pt}{1-qt}$$

$$\begin{aligned} Y \sim \text{NB}(r, p), Y = \underbrace{X + X + \dots + X}_{r \text{ terms (successes)}} \Rightarrow G_Y(t) \\ = G_{\underbrace{X+X+\dots+X}_{r \text{ terms}}}(t) = \underbrace{G_X(t) \times G_X(t) \times \dots \times G_X(t)}_{r \text{ factors}} \\ = \underbrace{\frac{pt}{1-qt} \times \frac{pt}{1-qt} \times \dots \times \frac{pt}{1-qt}}_{r \text{ factors}} = \left(\frac{pt}{1-qt} \right)^r \end{aligned}$$

$$\begin{aligned} 5 \quad a \quad \left. \begin{aligned} X_1 \sim \text{B}(n_1, p), G_{X_1}(t) &= (q + pt)^{n_1} \\ X_2 \sim \text{B}(n_2, p), G_{X_2}(t) &= (q + pt)^{n_2} \end{aligned} \right\} \Rightarrow G_{X_1+X_2}(t) \\ = G_{X_1}(t) \times G_{X_2}(t) = (q + pt)^{n_1} \times (q + pt)^{n_2} \\ = (q + pt)^{n_1+n_2} \Rightarrow X_1 + X_2 \sim \text{B}(n_1 + n_2, p) \end{aligned}$$

b 1st Base

For $n = 1$, the statement is trivial: $X_1 \sim \text{B}(n_1, p)$

For $n = 2$, in part **a** we proved that

$$X_1 + X_2 \sim \text{B}(n_1 + n_2, p)$$

2nd Assumption

Let's assume that for $n = k$ this statement is true:

$$\sum_{i=1}^k X_i \sim \text{B}\left(\sum_{i=1}^k n_i, p\right)$$

3rd Step

$$\text{For } n = k + 1, \sum_{i=1}^{k+1} X_i = \sum_{i=1}^k X_i + X_{k+1},$$

by the assumption the first term has a binomial distribution and by the given proposition in the question $X_{k+1} \sim \text{B}(n_{k+1}, p)$. Now we use the fact from part **a** that the sum of two binomial variables with the same probability p satisfies the following:

$$\begin{aligned} \left. \sum_{i=1}^k X_i \sim \text{B}\left(\sum_{i=1}^k n_i, p\right) \right\} \Rightarrow \sum_{i=1}^k X_i + X_{k+1} \sim \\ X_{k+1} \sim \text{B}(n_{k+1}, p) \\ \text{B}\left(\sum_{i=1}^k n_i + n_{k+1}, p\right) \Rightarrow \sum_{i=1}^{k+1} X_i \sim \text{B}\left(\sum_{i=1}^{k+1} n_i, p\right) \end{aligned}$$

4th Conclusion

$$\text{Now we can conclude that } Y = \sum_{i=1}^k X_i \sim \text{B}\left(\sum_{i=1}^k n_i, p\right)$$

for all the positive integer values of k .

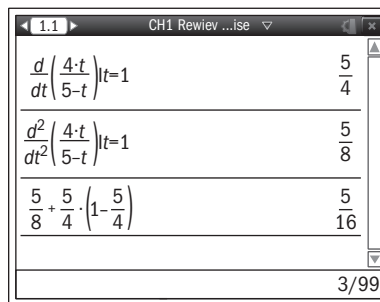
Review exercise

$$\begin{aligned} 1 \quad a \quad G(t) &= \frac{4t}{5-t} = \frac{4}{5}t \times \frac{1}{1-\frac{t}{5}} \\ &= \frac{4}{5}t \times \left(1 + \frac{1}{5}t + \frac{1}{25}t^2 + \frac{1}{125}t^3 + \dots\right) \\ &= \frac{4}{5}t + \frac{4}{25}t^2 + \frac{4}{125}t^3 + \frac{4}{625}t^4 + \dots \\ \Rightarrow P(1 \leq X \leq 4) &= \frac{4}{5} + \frac{4}{25} + \frac{4}{125} + \frac{4}{625} = \frac{624}{625} \end{aligned}$$

$$b \quad P(X \geq 2) = 1 - P(X \leq 1) = 1 - \left(\frac{4}{5} + \frac{4}{25}\right) = 1 - \frac{24}{25} = \frac{1}{25}$$

$$\begin{aligned} c \quad G(t) &= \frac{4t}{5-t}, G'(t) = \frac{4(5-t) + 4t}{(5-t)^2} \\ &= \frac{20}{(5-t)^2} \Rightarrow E(X) = G'(1) = \frac{20}{16} = \frac{5}{4} \end{aligned}$$

$$\begin{aligned} d \quad G'(t) &= 20(5-t)^{-2}, \\ G''(t) &= -40(5-t)^{-3} \times (-1) = 40(5-t)^{-3} \\ \Rightarrow \text{Var}(X) &= G''(1) + G'(1)(1 - G'(1)) \\ &= 40 \times 4^{-3} + \frac{5}{4} \left(1 - \frac{5}{4}\right) = \frac{5}{8} - \frac{5}{16} = \frac{5}{16} \end{aligned}$$



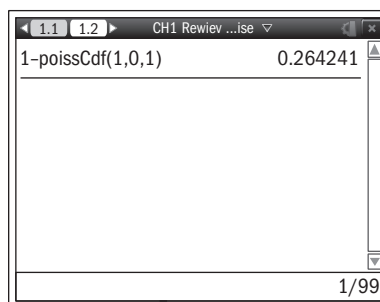
$$2 \quad a \quad G_X(t) = e^{0.6(t-1)}, G_Y(t) = e^{0.12(t-1)}, G_Z(t) = e^{0.28(t-1)}$$

$$\begin{aligned} b \quad G_{X+Y+Z}(t) &= G_X(t) \times G_Y(t) \times G_Z(t) \\ &= e^{0.6(t-1)} \times e^{0.12(t-1)} \times e^{0.28(t-1)} = e^{t-1} \end{aligned}$$

$$\begin{aligned} G_{X+Y+Z}(t) &= e^{t-1} \Rightarrow E(X + Y + Z) \\ &= G_{X+Y+Z}'(1) = e^{1-1} = e^0 = 1 \end{aligned}$$

The expected number of animals at the meadow at the given moment is 1.

$$\begin{aligned} c \quad G_{X+Y+Z}(t) &= e^{t-1} \Rightarrow X + Y + Z \sim \text{Po}(1) \\ P(X + Y + Z \geq 2) &= 1 - P(X + Y + Z \leq 1) = 0.264 \end{aligned}$$



3 a i $X \sim \text{Geo}(0.75) \Rightarrow P(X = 4) = 0.0117$

ii $Y \sim \text{NB}(2, 0.75) \Rightarrow P(Y = 4)$
 $= \binom{4-1}{2-1} 0.25^2 \times 0.75^2 = \frac{27}{256} = 0.105$

iii $P(X \geq 3) = 1 - P(X \leq 2) = 0.0625$

iv $Z \sim \text{NB}(6, 0.75)$

$$\Rightarrow P(Z \leq 10) = \binom{6-1}{6-1} 0.75^6$$

$$+ \binom{7-1}{6-1} 0.75^6 \times 0.25 + \dots$$

$$+ \binom{10-1}{6-1} 0.75^6 \times 0.25^4 = 0.922$$

b $E(Z) = \frac{6}{0.75} = 8$, the expected number of students needed to be selected if we need six involved in the programme is eight.

geomPdf(0.75,4)	0.011719
nCr(3,1)·(0.25) ² ·(0.75) ² →approxFraction(5.1)	$\frac{27}{256}$
1-geomCdf(0.75,2)	0.0625
	3/99

$(0.75)^6 \cdot \sum_{k=0}^4 \left(nCr(5+k,5) \cdot (0.25)^k \right)$	0.921873
$\frac{6}{0.75}$	8.
	5/99

4 a $F(2) = 1 \Rightarrow \frac{2^2}{a^2} = 1 \Rightarrow a^2 \Rightarrow 2^2 \Rightarrow a = 2$ or

~~$a = -2$~~ since a must be positive.

b $F'(x) = f(x) \Rightarrow f(x) = \begin{cases} \frac{2x}{2^2}, & 0 \leq x \leq 2 \\ 0, & \text{otherwise} \end{cases}$

$$= \begin{cases} \frac{x}{2}, & 0 \leq x \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

c Since $f(x)$ is increasing on the whole interval $[0, 2]$ the modal value of the variable X is 2.

5 a $G(1) = 1 \Rightarrow \frac{1}{a-b} = 1 \Rightarrow a-b=1 \Rightarrow b=a-1$

b $G'(t) = \frac{a-bt-t \times (-b)}{a-bt^2} = \frac{a}{(a-bt)^2}$
 $\Rightarrow E(X) = G'(1) = \frac{a}{(a-b)^2} = \frac{a}{1^2} = a$

c $G'(t) = a(a-bt)^{-2}$,
 $G''(t) = a \times (-2)(a-bt)^{-3} \times (-b) = \frac{2ab}{(a-bt)^3}$
 $\text{Var}(X) = G''(1) + G'(1)(1-G'(1))$
 $= \frac{2a(a-1)}{1^3} + a(1-a)$
 $= 2a^2 - 2a + a - a^2 = a^2 - a$

6 a $\left. \begin{aligned} X_1 &\sim \text{Po}(m), G_{X_1}(t) = e^{m(t-1)} \\ X_2 &\sim \text{Po}(m), G_{X_2}(t) = e^{m(t-1)} \end{aligned} \right\} \Rightarrow G_{X_1+X_2}(t)$
 $= G_{X_1}(t) \times G_{X_2}(t) = e^{m(t-1)} \times e^{m(t-1)} = e^{m(t-1)+m(t-1)}$
 $= e^{2m(t-1)} \Rightarrow X_1 + X_2 \sim \text{Po}(2m)$

b 1st Base

For $n=1$, the statement is trivial: $X_1 \sim \text{Po}(m)$

For $n=2$, in part a we proved that $X_1 + X_2 \sim \text{Po}(2m)$

2nd Assumption

Let's assume that for $n=k$ this statement is true:

$$\sum_{i=1}^k X_i \sim \text{Po}(km)$$

3rd Step

For $n=k+1$, $\sum_{i=1}^{k+1} X_i = \sum_{i=1}^k X_i + X_{k+1}$, by the

assumption the first term has a Poisson distribution with the coefficient km and by the given proposition in the question $X_{k+1} \sim \text{Po}(m)$. Now we use the fact that the coefficient of the sum of two Poisson variables is the sum of the coefficients of those two variables.

$$\left. \begin{aligned} \sum_{i=1}^k X_i &\sim \text{Po}(km) \\ X_{k+1} &\sim \text{Po}(m) \end{aligned} \right\} \Rightarrow \sum_{i=1}^{k+1} X_i + X_{k+1} \sim \text{Po}(km+m)$$

$$\Rightarrow \sum_{i=1}^{k+1} X_i \sim \text{Po}((k+1)m)$$

4th Conclusion

Now we can conclude that $Y = \sum_{k=1}^n X_i \sim \text{B}(nm)$

for all the positive integer values of n .

2

Expectation algebra and Central Limit Theorem

Skills check

$$1 \quad \left. \begin{aligned} np &= 2 \\ npq &= \frac{3}{2} \end{aligned} \right\} \Rightarrow 2q = \frac{3}{2} \Rightarrow q = \frac{3}{4},$$

$$p = \frac{1}{4}, n = 8 \Rightarrow X \sim B\left(8, \frac{1}{4}\right)$$

$$a \quad P(X = 2) = \binom{8}{2} \left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^6 = 0.311$$

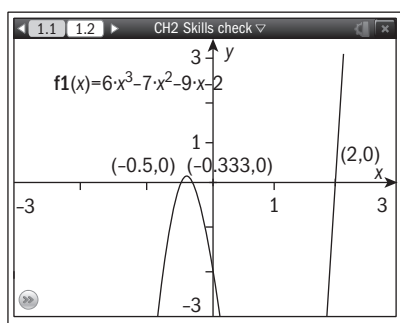
$$b \quad P(1 \leq X \leq 3) = \sum_{k=1}^3 \binom{8}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{8-k} = 0.786$$

binomPdf $\left(8, \frac{1}{4}, 2\right)$	0.311462
binomCdf $\left(8, \frac{1}{4}, 1, 3\right)$	0.786072
	2/99

$$2 \quad X \sim \text{Po}(m) \Rightarrow E(X) = \text{Var}(X)$$

$$7a^2 = 6a^3 - 9a - 2 \Rightarrow 6a^3 - 7a^2 - 9a - 2 = 0$$

$$\Rightarrow (2a + 1)(3a + 1)(a - 2) = 0$$



$$\Rightarrow a_1 = -\frac{1}{2}, a_2 = -\frac{1}{3}, a_3 = 2$$

We see that all three values of a give positive values of the parameter m :

$$m_1 = \frac{7}{4}, m_2 = \frac{7}{9}, m_3 = 28$$

Exercise 2A

$$1 \quad E(X) = 5.3, \text{Var}(X) = 1.2$$

$$a \quad E(3X) = 3 \times 5.3 = 15.9, \\ \text{Var}(3X) = 3^2 \times 1.2 = 10.8$$

$$b \quad E(X + 3) = 5.3 + 3 = 8.3, \text{Var}(X + 3) = 1.2$$

$$c \quad E(4X + 1) = 4 \times 5.3 + 1 = 22.2, \\ \text{Var}(4X + 1) = 4^2 \times 1.2 = 19.6$$

$$d \quad E(2X - 5) = 2 \times 5.3 - 5 = 5.6, \\ \text{Var}(2X - 5) = 2^2 \times 1.2 = 4.8$$

$$e \quad E(kX + p) = k \times 5.3 + p, \\ \text{Var}(kX + p) = k^2 \times 1.2 = 1.2k^2$$

$$2 \quad X \sim B\left(10, \frac{2}{5}\right) \Rightarrow E(X) = 10 \times \frac{2}{5} = 4,$$

$$\text{Var}(X) = 10 \times \frac{2}{5} \times \frac{3}{5} = \frac{12}{5} = 2.4$$

$$a \quad E(3X + 2) = 3 \times E(X) + 2 = 3 \times 4 + 2 = 14$$

$$b \quad \text{Var}(3X - 2) = 3^2 \times \text{Var}(X) = 9 \times \frac{12}{5} = \frac{108}{5} = 21.6$$

$$3 \quad Y \sim \text{Geo}\left(\frac{2}{3}\right) \Rightarrow E(Y) = \frac{1}{\frac{2}{3}} = \frac{3}{2} = 1.5,$$

$$\text{Var}(Y) = \frac{1}{\left(\frac{2}{3}\right)^2} = \frac{3}{4} = 0.75$$

$$E(2Y - 1) = 2 \times E(Y) - 1 = 2 \times \frac{3}{2} - 1 = 2,$$

$$\text{Var}(2Y - 1) = 2^2 \times \text{Var}(Y) = 4 \times \frac{3}{4} = 3$$

$$4 \quad Y \sim \text{Po}(2) \Rightarrow E(Y) = 2, \text{Var}(Y) = 2$$

$$a \quad E(3 - 2Y) = 3 - 2 \times E(Y) = 3 - 2 \times 2 = -1$$

$$b \quad \text{Var}(3 - 2Y) = 2^2 \times \text{Var}(Y) = 4 \times 2 = 8$$

$$5 \quad X \sim \text{NB}\left(8, \frac{1}{3}\right) \Rightarrow E(X) = \frac{8}{\frac{1}{3}} = 24,$$

$$\text{Var}(Y) = \frac{8 \times \frac{2}{3}}{\left(\frac{1}{3}\right)^2} = 48$$

$$a \quad E(2X - 3) = 2 \times E(X) - 3 = 2 \times 24 - 3 = 45$$

$$b \quad \text{Var}(2X - 11) = 2^2 \times \text{Var}(X) = 4 \times 48 = 192$$

$$6 \quad X \sim B(15, p) \Rightarrow E(X) = np \Rightarrow 15p = 6 \Rightarrow p = \frac{2}{5}$$

$$\text{Var}(X) = npq \Rightarrow \text{Var}(X) = 6 \times \frac{3}{5} = \frac{18}{5} = 3.6$$

$$\text{Var}(5X + 3) = 5^2 \times \frac{18}{5} = 90$$

$$7 \quad E(X) = \int_0^{\sqrt{6}} x \times \frac{1}{3} x dx = \frac{1}{3} \int_0^{\sqrt{6}} x^2 dx = \frac{1}{3} \left[\frac{x^3}{3} \right]_0^{\sqrt{6}} \\ = \frac{1}{9} 6\sqrt{6} = \frac{2}{3} \sqrt{6}$$

$$\text{Var}(X) = \int_0^{\sqrt{6}} x^2 \times \frac{1}{3} x dx - \left(\frac{2}{3} \sqrt{6} \right)^2 \\ = \frac{1}{3} \int_0^{\sqrt{6}} x^3 dx - \frac{8}{3} = \frac{1}{3} \left[\frac{x^4}{4} \right]_0^{\sqrt{6}} - \frac{8}{3} \\ = \frac{1}{12} \times 6^2 - \frac{8}{3} = 3 - \frac{8}{3} = \frac{1}{3}$$

$$E(3X + 2) = 3 \times \frac{2}{3} \sqrt{6} + 2 = 2\sqrt{6} + 2,$$

$$\text{Var}(3X + 2) = 3^2 \times \frac{1}{3} = 3$$

Exercise 2B

$$1 \quad a \quad E(X + Y) = 3 + (-5) = -2,$$

$$\text{Var}(X + Y) = 0.5 + 1.4 = 1.9$$

$$b \quad E(2Y - Z) = 2 \times (-5) - 12 = -22,$$

$$\text{Var}(2Y - Z) = 2^2 \times 1.4 + 2.8 = 8.4$$

$$c \quad E(2Z - 7X) = 2 \times 12 - 7 \times 3 = 3,$$

$$\text{Var}(2Z - 7X) = 2^2 \times 2.8 + 7^2 \times 0.5 = 35.7$$

$$d \quad E(X - Y + Z) = 3 - (-5) + 12 = 20,$$

$$\text{Var}(X - Y + Z) = 0.5 + 1.4 + 2.8 = 4.7$$

$$e \quad E(X + Y - Z) = 3 + (-5) - 12 = -14,$$

$$\text{Var}(X + Y - Z) = 0.5 + 1.4 + 2.8 = 4.7$$

$$f \quad E(3Z - 2X + 4Y) = 3 \times 12 - 2 \times 3 + 4 \times (-5) = 10,$$

$$\text{Var}(3Z - 2X + 4Y) = 3^2 \times 2.8 + 2^2 \times 0.5 + 4^2 \times 1.4 = 49.6$$

$$2 \quad \text{Var}(X) = 2 \Rightarrow X \sim \text{Po}(2) \Rightarrow E(X) = 2 \text{ and}$$

$$E(Y) = 5 \Rightarrow Y \sim \text{Po}(5) \Rightarrow \text{Var}(Y) = 5$$

$$a \quad E(3X + 5Y) = 3 \times 2 + 5 \times 5 = 31$$

$$b \quad \text{Var}(11Y - 7X) = 11^2 \times 5 + 7^2 \times 2 = 703$$

$$3 \quad E(X) = 9 \Rightarrow n_1 p = 9 \Rightarrow \text{Var}(X) = 9(1 - p)$$

$$E(Y) = 4 \Rightarrow n_2 p = 4 \Rightarrow \text{Var}(Y) = 4(1 - p)$$

$$\text{Var}(2X - 3Y) = 2^2 \times 9(1 - p) + 3^2 \times 4(1 - p) \\ = 72 - 72p$$

$$4 \quad E(X) = 8 \Rightarrow \frac{r_1}{p} = 8 \Rightarrow$$

$$\text{Var}(X) = 8 \times \frac{1 - p}{p} = \frac{8}{p} - 8$$

$$E(Y) = 12 \Rightarrow \frac{r_2}{p} = 12 \Rightarrow$$

$$\text{Var}(Y) = 12 \times \frac{1 - p}{p} = \frac{12}{p} - 12$$

$$\text{Var}(X - Y) = \frac{8}{p} - 8 + \frac{12}{p} - 12 = \frac{20}{p} - 20$$

$$5 \quad \text{Given the } n \text{ independent variables and their corresponding parameters}$$

$$X_n, E(X_n), \text{Var}(X_n), n \in \mathbb{Z}^+$$

$$1\text{st Base}$$

$$n = 1$$

$$X_1, E(X_1), \text{Var}(X_1) \Rightarrow E(aX_1) = aE(X_1),$$

$$\text{Var}(aX_1) = a^2 \text{Var}(X_1)$$

2nd Assumption

Let's assume that for $n = k$ the statement is true so the parameters of $\sum_{i=1}^k a_i X_i$ are calculated as follows:

$$E\left(\sum_{i=1}^k a_i X_i\right) = \sum_{i=1}^k a_i E(X_i) \text{ and}$$

$$\text{Var}\left(\sum_{i=1}^k a_i X_i\right) = \sum_{i=1}^k a_i^2 \text{Var}(X_i)$$

3rd Step

$$\text{For } n = k + 1, \sum_{i=1}^{k+1} X_i = \sum_{i=1}^k X_i + X_{k+1},$$

$$\begin{aligned} E\left(\sum_{i=1}^{k+1} a_i X_i\right) &= E\left(\sum_{i=1}^k a_i X_i + a_{k+1} X_{k+1}\right) \\ &= E\left(\sum_{i=1}^k a_i X_i\right) + E(a_{k+1} X_{k+1}) \\ &= \sum_{i=1}^k a_i E(X_i) + a_{k+1} E(X_{k+1}) = \sum_{i=1}^{k+1} a_i E(X_i) \end{aligned}$$

$$\begin{aligned} \text{Var}\left(\sum_{i=1}^{k+1} a_i X_i\right) &= \text{Var}\left(\sum_{i=1}^k a_i X_i + a_{k+1} X_{k+1}\right) \\ &= \sum_{i=1}^k a_i^2 \text{Var}(X_i) = a_{k+1}^2 \text{Var}(X_{k+1}) \\ &= \text{Var}\left(\sum_{i=1}^{k+1} a_i X_i\right) \end{aligned}$$

4th Conclusion

The parameters of a linear combination of n random variables can be calculated by:

$$E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(X_i),$$

$$\text{Var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \text{Var}(X_i)$$

Exercise 2C

- 1 Let's independently flip 6 unbiased coins and record the number of heads obtained.

a

$X = x$	0	1
$P\{X = x\}$	$\frac{1}{2}$	$\frac{1}{2}$

$$E(X) = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2},$$

$$\text{Var}(X) = 0^2 \times \frac{1}{2} + 1^2 \times \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$

- b Since we have 6 independent flips of the same coin we add six instances of the variable X .

- c Find the expected value and variance of the random variable Y .

$$E(Y) = 6 \times \frac{1}{2} = 3, \text{Var}(Y) = 6 \times \frac{1}{4} = \frac{3}{2}$$

d $E(Y) \pm 3 \times \sqrt{\text{Var}(Y)} = 3 \pm 3 \times \sqrt{\frac{3}{2}}$
 $= 3 \pm 3.67 \Rightarrow y \in [-0.67, 6.67]$

We notice that all the possible outcomes $\{0, 1, 2, 3, 4, 5, 6\}$ are covered by the 99.7% empirical rule.

- 2 Let's roll an unbiased die 4 times. Record the number of multiples of 3 obtained.

a

$X = x$	0	1
$P\{X = x\}$	$\frac{2}{3}$	$\frac{1}{3}$

$$E(X) = 0 \times \frac{2}{3} + 1 \times \frac{1}{3} = \frac{1}{3},$$

$$\text{Var}(X) = 0^2 \times \frac{2}{3} + 1^2 \times \frac{1}{3} - \left(\frac{1}{3}\right)^2 = \frac{1}{3} - \frac{1}{9} = \frac{2}{9}$$

- b Since we have 4 independent rolls of the same die we add four instances of the variable X .

c $E(Y) = 4 \times \frac{1}{3} = \frac{4}{3}, \text{Var}(Y) = 4 \times \frac{2}{9} = \frac{8}{9}$

- 3 a $E(X + X + X) = 3 \times E(X) = 3 \times 3 = 9,$

$$\text{Var}(X + X + X) = 3 \times \text{Var}(X) = 3 \times 4 = 12$$

- b $E(3X) = 3 \times E(X) = 3 \times 3 = 9,$

$$\text{Var}(3X) = 3^2 \times \text{Var}(X) = 9 \times 4 = 36$$

- c $\text{Var}(X + X + X + 3X) = 3 \times \text{Var}(X)$

$$+ 3^2 \times \text{Var}(X) = 12 \text{Var}(X) = 12 \times 4 = 48$$

$$\text{Var}(6X) = 6^2 \text{Var}(X) = 36 \times 4 = 144 \Rightarrow$$

$$\text{Var}(X + X + X + 3X) \neq \text{Var}(6X)$$

- 4 a $E(X + X + X + X + X) = 5 \times E(X) = 5 \times 2 = 10$

$$\text{Var}(X + X + X + X + X) = 5 \times \text{Var}(X) = 5 \times 1 = 5$$

- b $E(Y + Y + Y) = 3 \times E(Y) = 3 \times 5 = 15$

$$\text{Var}(Y + Y + Y) = 3 \times \text{Var}(Y) = 3 \times 3 = 9$$

- c $E(X + X + X + X + X + Y + Y + Y)$

$$= E(X + X + X + X + X) + E(Y + Y + Y)$$

$$= 10 + 15 = 25$$

$$\begin{aligned}\text{Var}(X + X + X + X + X + Y + Y + Y) \\&= \text{Var}(X + X + X + X + X) + \text{Var}(Y + Y + Y) \\&= 5 + 9 = 14\end{aligned}$$

5 a

$X = x_i$	1	2	3	4
$P\{X = x_i\}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

$$E(X) = \frac{1}{4} \times (1 + 2 + 3 + 4) = \frac{1}{4} \times 10 = \frac{5}{2}$$

$Y = y_i$	1	2	3
$P\{Y = y_i\}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

$$E(Y) = \frac{1}{2} \times 1 + \frac{1}{3} \times 2 + \frac{1}{6} \times 3 = \frac{5}{3}$$

b 1: $P(\{1, 1\}) = \frac{1}{4} \times \frac{1}{2} = \frac{1}{8}$

2: $P(\{1, 2\} \text{ or } \{2, 1\}) = \frac{1}{4} \times \frac{1}{3} + \frac{1}{4} \times \frac{1}{2} = \frac{5}{24}$

3: $(1, 3) \text{ or } (3, 1) \rightarrow \frac{1}{4} \times \frac{1}{6} + \frac{1}{4} \times \frac{1}{2} = \frac{1}{6}$

4: $P(\{2, 2\} \text{ or } \{4, 1\}) = \frac{1}{4} \times \frac{1}{3} + \frac{1}{4} \times \frac{1}{2} = \frac{5}{24}$

6: $P(\{2, 3\} \text{ or } \{3, 2\}) = \frac{1}{4} \times \frac{1}{6} + \frac{1}{4} \times \frac{1}{3} = \frac{1}{8}$

8: $P(\{4, 2\}) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12}$

9: $P(\{3, 3\}) = \frac{1}{4} \times \frac{1}{6} = \frac{1}{24}$

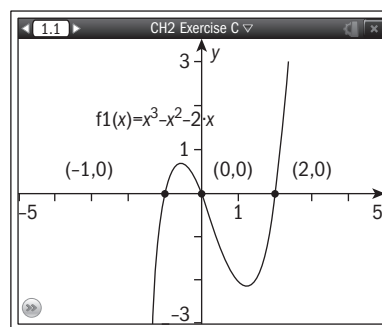
12: $P(\{4, 3\}) = \frac{1}{4} \times \frac{1}{6} = \frac{1}{24}$

$Z = z_i$	1	2	3	4	6	8	9	12
$P(Z = z_i)$	$\frac{1}{8}$	$\frac{5}{24}$	$\frac{1}{6}$	$\frac{5}{24}$	$\frac{1}{8}$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{24}$

$$\begin{aligned}E(Z) &= 1 \times \frac{1}{8} + 2 \times \frac{5}{24} + 3 \times \frac{1}{6} \\&\quad + 4 \times \frac{5}{24} + 6 \times \frac{1}{8} + 8 \times \frac{1}{12} \\&\quad + 9 \times \frac{1}{24} + 12 \times \frac{1}{24} = \frac{25}{6}\end{aligned}$$

$$E(X) \times E(Y) = \frac{5}{2} \times \frac{5}{3} = \frac{25}{6} = E(Z) = E(XY)$$

6 $E(X) \times E(Y) = E(XY) \Rightarrow \mu \times (\mu + 2) = \mu^3$
 $\Rightarrow \mu^3 - \mu^2 - 2\mu = 0$
 $\Rightarrow \mu(\mu + 1)(\mu - 2) = 0$
 $\Rightarrow \mu \neq 0 \text{ or } \mu \neq -1 \text{ or } \mu = 2.$



Exercise 2D

1 a $\mu = 1 - (-2) - 3 = 0, \sigma^2 = 0.16 + 0.25 + 1.21 = 1.62$

$$Y - Z - W \sim N(0, 1.62)$$

$$\Rightarrow P(Y - Z - W < 0) = 0.5$$

b $\mu = 0 + 1 + (-2) + 3 = 2,$

$$\sigma^2 = 1 + 0.16 + 0.25 + 1.21 = 2.62$$

$$X + Y + Z + W \sim N(2, 2.62)$$

$$\Rightarrow P(X + Y + Z + W > 0) = 0.892$$

c $\mu = 3 \times 0 + 1 - (-2) - 3 = 0,$

$$\sigma^2 = 3^2 \times 1 + 0.16 + 0.25 + 1.21 = 10.62$$

$$3X + Y - Z - W \sim N(0, 10.62)$$

$$\Rightarrow P(3X + Y > Z + W)$$

$$= P(3X + Y - Z - W > 0) = 0.5$$

d $\mu = 0 - 2 \times 1 - 3 \times (-2) - 3 = -1,$

$$\sigma^2 = 1 + 2^2 \times 0.16 + 3^2 \times 0.25 + 1.21 = 5.1$$

$$X - 2Y - 3Z - W \sim N(1, 5.1)$$

$$\Rightarrow P(X - 3Z \leq 2Y + W)$$

$$= P(X - 2Y - 3Z - W \leq 0) = 0.329$$

e $\mu = 0 - 4 \times (-2) - 3 \times 0 + 2(-2) = 4,$

$$\sigma^2 = 1 + 4^2 \times 0.25 + 3^2 \times 1 + 2^2 \times 0.25 = 15$$

$$X - 4Z - 3X + 2Z \sim N(4, 15)$$

$$\Rightarrow P(X - 4Z \leq 3X - 2Z)$$

$$= P(X - 4Z - 3X + 2Z \leq 0) = 0.151$$

f $\mu = -3 \times 1 - 2 \times 3 = -9$,
 $\sigma^2 = 1.21 + 0.16 + 2^2 \times 0.16 + 3^2 \times 1.21 = 12.9$
 $W - Y - 2Y - 3W \sim N(-9, 12.9)$
 $\Rightarrow P(W - Y \leq 2Y + 3W)$
 $= P(W - Y - 2Y - 3W \leq 0) = 0.994$

normCdf(-1000,0,0, $\sqrt{1.62}$)	0.5
normCdf(0,1000,2, $\sqrt{2.62}$)	0.891697
normCdf(-1000,0,-1, $\sqrt{5.1}$)	0.671047
normCdf(-1000,0,2, $\sqrt{14.25}$)	0.298121
normCdf(-1000,0,-9, $\sqrt{17.21}$)	0.984976
5/99	

2 a $X \sim N(240, 20^2)$, $Y \sim N(730, 50^2)$
 $\Rightarrow Y - 3X \sim N(10, 50^2 + 3^2 \times 20^2)$
 $\Rightarrow P(Y - 3X \geq 0) = 0.551$

b $X \sim N(240, 20^2)$, $Y \sim N(730, 50^2)$
 $\Rightarrow 2Y + 4X \sim N(2420, 2 \times 50^2 + 4 \times 20^2)$
 $\Rightarrow P(2Y + 4X \geq 2500) = 0.162$

3 a $L \sim N(35, 5^2) \Rightarrow P(L < 30) = 0.159$

b $V \sim N(45, 8^2) \Rightarrow$
 $W = +V + V + V + V + V \sim N(5 \times 45, 5 \times 8^2)$
 $= N(225, (8\sqrt{5})^2)$
 $P(W > 240) = 0.201$

c $4L - 3V \sim N(4 \times 35 - 3 \times 45, 4^2 \times 5^2 + 3^2 \times 8^2)$
 $= N(5, (4\sqrt{61})^2) \Rightarrow P(4L - 3V > 0) = 0.564$

normCdf(-9.999,30,35,5)	0.158655
normCdf(240,9.999,225,8 $\sqrt{5}$)	0.200868
normCdf(0,9.999,5,4 $\sqrt{61}$)	0.563578
3/99	

4 a $X \sim N(225, 12^2) \Rightarrow P(X > 250) = 0.0186$

b $X + X + X + X \sim N(4 \times 225, 4 \times 12^2) = N(900, 24^2)$
 $\Rightarrow P(X + X + X + X > 1000) = 0.0000155$

c $Y \sim N(60, 2^2) \Rightarrow 4Y \sim N(4 \times 60, 4^2 \times 2^2)$
 $= N(240, 8^2)$
 $X - 4Y \sim N(225 - 240, 12^2 + 8^2)$
 $= N(-15, (4\sqrt{13})^2)$
 $\Rightarrow P(X - 4Y > 0) = 0.149$

d $Y \sim N(60, 2^2) \Rightarrow$
 $Y + Y + Y + Y \sim N(4 \times 60, 4 \times 2^2)$
 $= N(240, 4^2)$
 $X - Y - Y - Y - Y \sim N(225 - 240, 12^2 + 4^2)$
 $= N(-15, (4\sqrt{10})^2)$
 $\Rightarrow P(X - Y - Y - Y - Y > 0) = 0.118$

normCdf(250,9.999,225,12)	0.01861
normCdf(1000,9.999,900,24)	0.000015
1.5463523972415E-5	0.000015
normCdf(0,9.999,-15,4 $\sqrt{13}$)	0.149155
normCdf(0,9.999,-15,4 $\sqrt{10}$)	0.11784
5/99	

Exercise 2E

1 a $\bar{X} \sim N\left(1.5, \left(\frac{2}{\sqrt{10}}\right)^2\right) \Rightarrow P(1 \leq \bar{X} \leq 3) = 0.777$

b $\bar{X} \sim N\left(5, \left(\frac{3}{\sqrt{7}}\right)^2\right) \Rightarrow P(|\bar{X} - 5| \leq 3)$
 $= P(2 \leq \bar{X} \leq 8) = 0.992$

c $\bar{X} \sim N\left(-0.2, \left(\frac{4}{\sqrt{22}}\right)^2\right) \Rightarrow P(|\bar{X}| \geq 0.8)$
 $= 1 - P(-0.8 \leq \bar{X} \leq 0.8) = 0.639$

normCdf(1,3,1.5, $\frac{2}{\sqrt{10}}$)	0.776549
normCdf(2,8,5, $\frac{3}{\sqrt{7}}$)	0.991849
normCdf(-0.8,0.8,-0.2, $\frac{4}{\sqrt{22}}$)	0.63867
3/99	

2 $X \sim N(5.5, 1.2^2) \Rightarrow \bar{X} \sim N\left(5.5, \left(\frac{1.2}{\sqrt{15}}\right)^2\right)$
 $\Rightarrow P(\bar{X} < 5) = 0.0533$

$$3 \quad X \sim N(35, 6^2) \Rightarrow \bar{X} \sim N\left(35, \left(\frac{6}{\sqrt{5}}\right)^2\right)$$

$$a \quad P(\bar{X} \geq 37) = 0.228$$

$$b \quad X + X + X + X + X \sim N(5 \times 35, 5 \times 6^2) \\ = N(175, (\sqrt{5} \times 6)^2) \\ \Rightarrow P(X + X + X + X + X > 180) = 0.355$$

normCdf(5,9.E999,5.5, $\frac{1.2}{\sqrt{15}}$)	0.946708
normCdf(37,9.E999,35, $\frac{6}{\sqrt{5}}$)	0.228028
normCdf(180,9.E999,175, $\sqrt{5} \cdot 6$)	0.354694
3/99	

$$4 \quad a \quad X \sim N(62.5, 18.5^2) \Rightarrow \bar{X} \sim N\left(62.5, \left(\frac{18.5}{\sqrt{12}}\right)^2\right)$$

$$P(\bar{X} \leq 70) = 0.920$$

$$b \quad T = \underbrace{X + X + \dots + X}_{12 \text{ terms}} \sim N(12 \times 62.5, 12 \times 18.5^2) \\ = N(750, (\sqrt{4107})^2) \\ \Rightarrow P(T \leq 800) = 0.782$$

normCdf(-9.E999,70,62.5, $\frac{18.5}{\sqrt{12}}$)	0.919895
normCdf(-9.E999,800,12 \cdot 62.5, $\sqrt{12} \cdot 18.5$)	0.782364
normCdf(-9.E999,800,750, $\sqrt{4107}$)	0.782364
3/99	

Exercise 2F

$$1 \quad a \quad \bar{X} \sim N\left(2, \left(\frac{3}{\sqrt{30}}\right)^2\right) \Rightarrow P(1.5 \leq \bar{X} \leq 2.5) = 0.639$$

$$b \quad \bar{X} \sim N\left(1.3, \left(\frac{0.2}{\sqrt{50}}\right)^2\right) \Rightarrow P(1.25 \leq \bar{X} \leq 1.35) = 0.923$$

$$c \quad \bar{X} \sim N\left(-0.5, \left(\frac{1}{10}\right)^2\right) \Rightarrow P(\bar{X} \geq -0.48) = 0.421$$

$$d \quad \bar{X} \sim N\left(400, \left(\sqrt{\frac{234}{85}}\right)^2\right) \Rightarrow P(\bar{X} < 397) = 0.0353$$

$$e \quad \bar{X} \sim N\left(1, \left(\sqrt{\frac{6}{40}}\right)^2\right) \Rightarrow P\left(|\bar{X}| < \frac{1}{2}\right)$$

$$= P\left(-\frac{1}{2} < \bar{X} < \frac{1}{2}\right) = 0.0983$$

$$f \quad \bar{X} \sim N\left(1.9, \left(\frac{6}{\sqrt{120}}\right)^2\right) \Rightarrow P\left(|\bar{X} - 2| \geq \frac{1}{5}\right)$$

$$= P(1.8 < \bar{X} < 2.2) = 0.280$$

normCdf(1.5,2.5,2, $\frac{3}{\sqrt{30}}$)	0.63869
normCdf(1.25,1.35,1.3, $\frac{0.2}{\sqrt{50}}$)	0.9229
normCdf(-0.48,9.E999,-0.5,0.1)	0.42074
3/99	

normCdf(-9.E999,397,400, $\sqrt{\frac{234}{85}}$)	0.035295
normCdf(- $\frac{1}{2}, \frac{1}{2}, 1, \sqrt{\frac{6}{40}}$)	0.098299
normCdf(1.8,2.2,1.9, $\frac{6}{\sqrt{120}}$)	0.280493
6/99	

$$2 \quad \bar{X} \sim N\left(30000, \left(\frac{8000}{\sqrt{15}}\right)^2\right)$$

$$P(\bar{X} \leq 32000) = 0.834$$

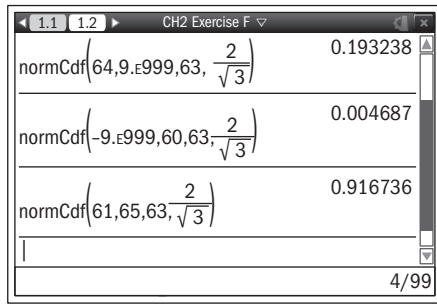
normCdf(-9.E999,32000,30000, $\frac{8000}{\sqrt{15}}$)	0.833539
1/99	

$$3 \quad X \sim N(63, 2^2) \Rightarrow \bar{X} \sim N\left(63, \left(\frac{2}{\sqrt{3}}\right)^2\right)$$

$$a \quad P(\bar{X} > 64) = 0.193$$

$$b \quad P(\bar{X} < 60) = 0.00469$$

c $P(61 \leq \bar{X} \leq 65) = 0.917$



4 $X \sim N(6, 4^2) \Rightarrow \bar{X} \sim N\left(6, \left(\frac{4}{\sqrt{n}}\right)^2\right)$

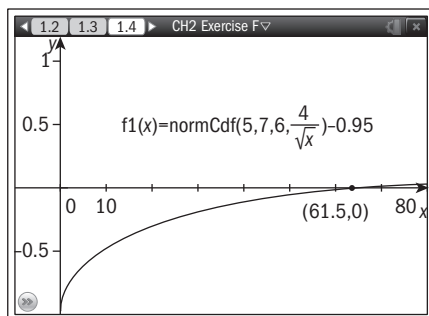
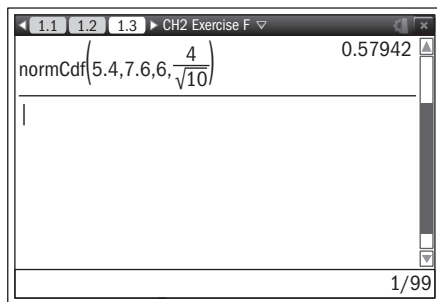
a $\bar{X} \sim N\left(6, \left(\frac{4}{\sqrt{10}}\right)^2\right) \Rightarrow P(5.4 \leq \bar{X} \leq 7.6) = 0.579$

b $P(5 \leq \bar{X} \leq 7) = 0.95 \Rightarrow P\left(\frac{5-6}{\frac{4}{\sqrt{n}}} \leq Z \leq \frac{7-6}{\frac{4}{\sqrt{n}}}\right)$

$= 0.95 \Rightarrow P\left(Z \leq \frac{\sqrt{n}}{4}\right) = 0.975$

$\Rightarrow \frac{\sqrt{n}}{4} = \Phi^{-1}(0.975) \Rightarrow \sqrt{n} = 4 \times 1.95996$

$\Rightarrow \sqrt{n} = 7.839856 \Rightarrow n = 61.4633 \approx 62$



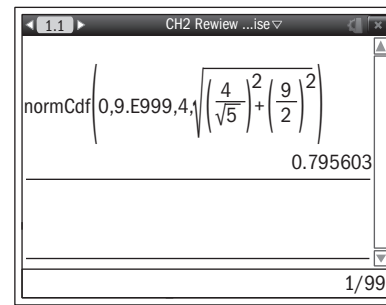
Review exercise

1 $B \sim N(202, 4^2) \Rightarrow \bar{B} \sim N\left(202, \left(\frac{4}{\sqrt{5}}\right)^2\right)$

$S \sim N(198, 9^2) \Rightarrow \bar{S} \sim N\left(198, \left(\frac{9}{\sqrt{4}}\right)^2\right)$

$\bar{B} - \bar{S} \sim N\left(4, \left(\frac{4}{\sqrt{5}}\right)^2 + \left(\frac{9}{\sqrt{4}}\right)^2\right)$

$\Rightarrow P(\bar{B} > \bar{S}) = P(\bar{B} - \bar{S} > 0) = 0.796$



2 a $\text{Var}(2X) = (E(X))^2 - 5 \Rightarrow 2^2 \text{Var}(X) = (E(X))^2 - 5$
 $\Rightarrow 4m = m^2 - 5 \Rightarrow m^2 - 4m - 5 = 0$

$\Rightarrow (m-5)(m+1) = 0 \Rightarrow m = 5$ or $m = -1$.
 Variance is always positive.

b $P(X \geq 6) = 1 - P(X \leq 5) = 0.384$

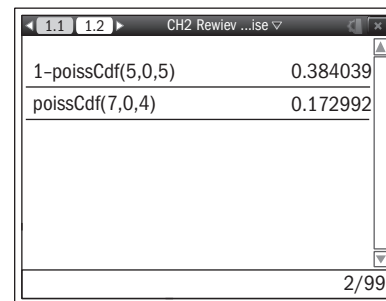
c $\text{Var}(3Y) = 18 \Rightarrow 3^2 \text{Var}(Y) = 18$

$\Rightarrow \text{Var}(Y) = 2 \Rightarrow Y \sim \text{Po}(2)$

$\Rightarrow X + Y \sim \text{Po}(5 + 2)$

$\Rightarrow X + Y \sim \text{Po}(7) \Rightarrow P(X + Y < 5)$

$= P(X + Y \leq 4) = 0.173$



d $E(Z) = E(3X - 4Y) = 3 \times 5 - 4 \times 2 = 7$

$\text{Var}(Z) = \text{Var}(3X - 4Y) = 3^2 \times 5 + 4^2 \times 2 = 77$

e The random variable Z has no Poisson distribution since $7 = E(Z) \neq \text{Var}(Z) = 77$.

3 a $X \sim N(400, 20^2) \Rightarrow P(X > 450) = 0.00621$

b $Y \sim N(350, 15^2) \Rightarrow Y + Y \sim N(2 \times 350, 2 \times 15^2)$

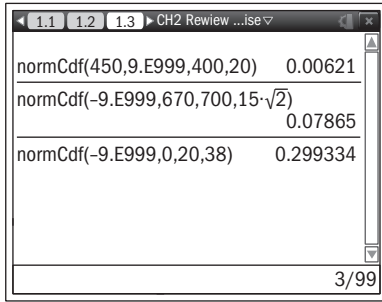
$\Rightarrow P(Y + Y < 670) = 0.0787$

c $Z \sim N(320, 12^2)$

$\Rightarrow X + Z \sim N(400 + 320, 20^2 + 12^2)$

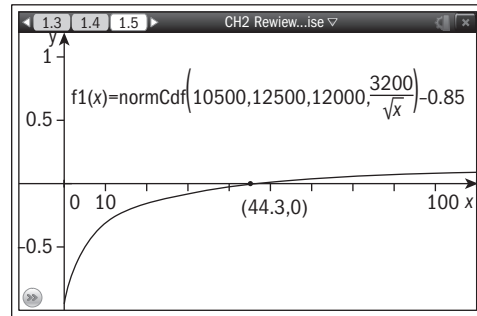
$= N(720, (4\sqrt{34})^2)$

d $X + Z - 2Y \sim N(720 - 2 \times 350, (4\sqrt{34})^2 + 2^2 \times 15^2)$
 $= N(20, 38^2)$
 $\Rightarrow P(X + Z - 2Y < 0) = 0.299$



c $\bar{X} \sim N\left(12000, \left(\frac{3200}{\sqrt{n}}\right)^2\right)$

$\Rightarrow P(1050 < \bar{X} < 12500) = 0.85 \Rightarrow n = 44.3 \approx 45$



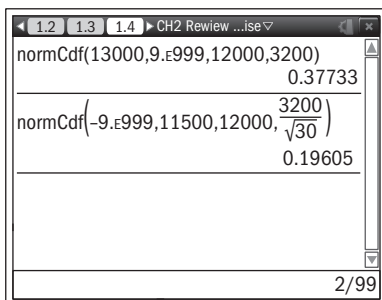
4 a $X \sim B(n, p) \Rightarrow \text{Var}(X) = 6 \Rightarrow npq = 6$
 $n = 27 \Rightarrow 27 \times pq = 6 \Rightarrow p(1 - p) = \frac{2}{9}$
 $\Rightarrow 9p^2 - 9p + 2 = 0 \Rightarrow (3p - 1)(3p - 2) = 0$
 $p = \frac{1}{3} \text{ or } p = \frac{2}{3}$
 $E(3X - 7) = 3E(X) - 7 = 3np - 7$

$\Rightarrow \begin{cases} E(3X - 7) = 3 \times 27 \times \frac{1}{3} - 7 = 20 \\ E(3X - 7) = 3 \times 27 \times \frac{2}{3} - 7 = 47 \end{cases}$

b $Z \sim \text{Po}(m) \Rightarrow (\text{Var}(Z))^2 = E(Z) + 12$
 $\Rightarrow m^2 - m - 12 = 0 \Rightarrow (m - 4)(m + 3) = 0$
 $m = 4 \text{ or } m = -3$ Variance is always positive.
 $\text{Var}(5 + 2Z) = 2^2 \text{Var}(Z) = 4 \times 4 = 16$

5 a $X \sim N(12000, 3200^2) \Rightarrow P(X > 13000) = 0.377$

b $\bar{X} \sim N\left(12000, \left(\frac{3200}{\sqrt{30}}\right)^2\right) \Rightarrow P(\bar{X} < 11500) = 0.196$



6 a $E(X - E(X))^2 = E(X^2 - 2XE(X) + (E(X))^2)$
 $= E(X^2) - 2E(X) \times E(X) + (E(X))^2$
 $= E(X^2) - 2(E(X))^2 + (E(X))^2$
 $= E(X^2) - \underbrace{(E(X))^2}_{\geq 0} \geq E(X^2)$

b $X \sim \text{Geo}(p)$ and $Y \sim \text{Geo}(q)$, where $p + q = 1$

$\Rightarrow E(X + Y) = E(X) + E(Y) = \frac{1}{p} + \frac{1}{q}$
 $= \frac{q + p}{pq} = \frac{1}{pq}$

$\Rightarrow \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

$= \frac{q}{p^2} + \frac{p}{q^2} = \frac{q^3 + p^3}{p^2q^2}$
 $= \frac{(q + p)(q^2 - pq + p^2)}{p^2q^2}$

$= \frac{(q + p)^2 - 3pq}{p^2q^2} = \frac{1 - 3pq}{p^2q^2}$

$E(X + Y)(E(X + Y) - 3)$

$= \frac{1}{pq} \left(\frac{1}{pq} - 3 \right) = \frac{1}{p^2q^2} - \frac{3}{pq}$

$= \frac{1 - 3pq}{p^2q^2} = \text{Var}(X + Y) \quad \text{Q.E.D.}$

3

Exploring statistical analysis methods

Skills check

1

CH3 Skills check				
A	B	C	D	
				=OneVar(a
1	1	22	Title	One-Var...
2	3	37	$\bar{x}$	3.61818
3	5	46	Σx	398.
4	7	5	Σx^2	1750.
5			$s_x := s_{n-1}$	1.68633
D2	=3.6181818181818			

CH3 Skills check				
A	B	C	D	E
				=OneVar(a
5		$s_x := s_{n-1}$	1.68633	
6		$\sigma_x := \sigma_{n-1}$	1.67865	2.81785
7		n	110.	
8		MinX	1.	
9		Q1X	3.	
E6	=d6 ²			

$$\bar{x} = 3.62, \sigma^2 = 1.67865^2 = 2.82$$

2 a $X \sim B\left(5, \frac{1}{2}\right) \Rightarrow P(X = 5)$

$$= \binom{5}{5} \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^5 = \frac{1}{32} = 0.03125$$

b $X \sim B\left(10, \frac{1}{5}\right) \Rightarrow P(3 \leq X < 8)$

$$= P(3 \leq X \leq 7) = \sum_{k=3}^7 \binom{10}{k} \left(\frac{4}{5}\right)^{10-k} \left(\frac{1}{5}\right)^k$$

$$= 0.322$$

CH3 Skills check				
A	B	C	D	
				binomPdf(5, 1/2, 5)
				0.03125
				binomCdf(8, 1/3, 5)
				0.980338
				$\sum_{k=0}^5 \binom{nCr(8,k) \cdot \left(\frac{2}{3}\right)^{8-k} \cdot \left(\frac{1}{3}\right)^k}{2187}$
				2144
				2187
				3/99

CH3 Skills check				
A	B	C	D	
				$\sum_{k=0}^5 \binom{nCr(8,k) \cdot \left(\frac{2}{3}\right)^{8-k} \cdot \left(\frac{1}{3}\right)^k}{2187}$
				0.640537
				1-binomCdf(12, 3/7, 4)
				0.640537
				binomCdf(10, 1/5, 7)-binomCdf(10, 1/5, 2)
				0.322123
				5/99

3 a $Y \sim \text{Po}(0.4) \Rightarrow P(Y = 0) = 0.670$

b $Y \sim \text{Po}(7) \Rightarrow P(3 \leq X < 8)$
 $= P(X \leq 7) - P(X \leq 2) = 0.569$

CH3 Skills check				
A	B	C	D	
				poissPdf(0.4, 0)
				0.67032
				poissCdf(3, 5)
				0.916082
				1-poissCdf(4, 7, 4)
				0.505391
				poissCdf(7, 7)-poissCdf(7, 2)
				0.569078
				4/99

Exercise 3A

- 1 a $\{2, 4, 6, 8, 10, 12, 14, 16, 18, 20\} \Rightarrow \bar{x} = 11, s^2 = 33$
 b $\{21, 24, 36, 28, 30, 22, 25, 26, 38, 32, 34, 29, 37, 33, 32, 34, 29, 37, 33, 31, 30\} \Rightarrow \bar{x} = 29.75, s^2 = 25.3$
 c $\{1, 4, 7, 10, \dots, 133\} \Rightarrow \bar{x} = 67, s^2 = 1518$

CH3 Exercise A				
A	B	C	D	
				mean(seq(2-x,x,1,10))
				11
				varPop(seq(2-x,x,1,10))
				33
				mean({21,24,36,28,30,22,25,26,38,32,34,29,37,33,32,34,29,37,33,31,30})
				119
				4
				varPop({21,24,36,28,30,22,25,26,38,32,34,29,37,33,32,34,29,37,33,31,30})
				405
				16
				3/6

CH3 Exercise A				
A	B	C	D	
				119
				4
				varPop({21,24,36,28,30,22,25,26,38,32,34,29,37,33,32,34,29,37,33,31,30})
				405
				16
				mean(seq(x,x,1,133,3))
				67
				varPop(seq(x,x,1,133,3))
				1518
				6/99

2 $\bar{x} = 0.65, s = 0.864$

CH3 Exercise A				
A	B	C	D	
				=OneVar(a
1	22		Title	One-Var...
2	12		$\bar{x}$	0.65
3	4		Σx	26.
4	2		Σx^2	46.
5			$s_x := s_{n-1}$	0.863802
E5	=0.86380197160799			

3 a $\bar{x} = 33.7$

b $s = 23.8$

	1.1	1.2	1.3	CH3 Exercise A
1	6		Title	One-Var...
2	13		$\bar{x}$	33.7143
3	26		Σx	2360.
4	17		Σx^2	118550.
5	8		$s_x := s_{n-1}x$	23.7695
E5	=23.769510891752			

Exercise 3B

1 a 99% CI [4.19, 5.81]

	1.1	CH3 Exercise B
zInterval 1.5,10,0.99: stat.results		
"Title"	"z Interval"	
"CLower"	4.18545	
"CUpper"	5.81455	
" $\bar{x}$ "	5.	
"ME"	0.814549	
"n"	10.	
" σ "	1.	
		1/99

b 95% CI [-12.8, -9.20]

	1.1	CH3 Exercise B
zInterval 4.3,-11,22,0.95: stat.results		
"Title"	"z Interval"	
"CLower"	-12.7968	
"CUpper"	-9.20318	
" $\bar{x}$ "	-11.	
"ME"	1.79682	
"n"	22.	
" σ "	4.3	
		2/99

c 90% CI [2819, 2889]

	1.1	CH3 Exercise B
zInterval 327,2854,230,0.9: stat.results		
"Title"	"z Interval"	
"CLower"	2818.53	
"CUpper"	2889.47	
" $\bar{x}$ "	2854.	
"ME"	35.4659	
"n"	230.	
" σ "	327.	
		3/99

2 a 90% CI [3.90, 6.10]

	1.1	1.2	CH3 Exercise B
zInterval 2,{1,2,3,4,5,6,7,8,9},1,0.9: stat.res			
"Title"	"z Interval"		
"CLower"	3.90343		
"CUpper"	6.09657		
" $\bar{x}$ "	5.		
"ME"	1.09657		
"sx := s _{n-1} x"	2.73861		
"n"	9.		
" σ "	2.		
			3/99

b 99% CI [0.104, 0.167]

	1.1	1.2	CH3 Exercise B
zInterval 0.04,{0.1,0.12,0.15,0.18,0.13,0.12}			
"Title"	"z Interval"		
"CLower"	0.104389		
"CUpper"	0.16652		
" $\bar{x}$ "	0.135455		
"ME"	0.031066		
"sx := s _{n-1} x"	0.035317		
"n"	11.		
" σ "	0.04		
			3/99

c 95% CI [320, 325]

	1.1	1.2	CH3 Exercise B
zInterval 4.5,{321,325,330,324,325,326,317}			
"Title"	"z Interval"		
"CLower"	320.5		
"CUpper"	325.214		
" $\bar{x}$ "	322.857		
"ME"	2.3572		
"sx := s _{n-1} x"	5.5727		
"n"	14.		
" σ "	4.5		
			3/99

3 $|\bar{x} - \mu| \leq 2 \Rightarrow |\mu - \bar{x}| \leq 2 \Rightarrow -2 \leq \mu - \bar{x} \leq 2$

$$\Rightarrow \bar{x} - 2 \leq \mu \leq \bar{x} + 2$$

$$\sigma = 5.5 \Rightarrow 1.960 \times \frac{5.5}{\sqrt{n}} = 2$$

$$\Rightarrow \sqrt{n} = 5.39 \Rightarrow n = 29.052$$

We should take at least 30 elements.

4 95% CI for the weight if elementines is [72.3 g, 77.2 g]

	1.1	1.2	1.3	CH3 Exercise B
zInterval 3.5,{70,75,77,71,68,80,85,72},1,0				
"Title"	"z Interval"			
"CLower"	72.3247			
"CUpper"	77.1753			
" $\bar{x}$ "	74.75			
"ME"	2.42533			
"sx := s _{n-1} x"	5.70088			
"n"	8.			
" σ "	3.5			
				3/99

5 a $\bar{x} = \frac{14.2 + 17.4}{2} = 15.8$

b $\sigma = 3 \Rightarrow 1.645 \times \frac{3}{\sqrt{n}} = 15.8 - 14.2$

$$\Rightarrow \sqrt{n} = 3.08 \Rightarrow n = 9.51$$

The sample size should be 10.

Investigation

Let's consider samples of different sizes: all have the mean value $\bar{x} = 100$ and they come from a population with $\sigma^2 = 64$.

- a i** $n = 10 \Rightarrow 95\% \text{ CI } [95.0, 105.0]$

zInterval 8,100,10,0.95: stat.results	
"Title"	"z Interval"
"CLower"	95.0416
"CUpper"	104.958
" $\bar{x}$ "	100.
"ME"	4.95836
"n"	10.
" σ "	8.

- ii** $n = 25 \Rightarrow 95\% \text{ CI } [96.9, 103.1]$

zInterval 8,100,25,0.95: stat.results	
"Title"	"z Interval"
"CLower"	96.8641
"CUpper"	103.136
" $\bar{x}$ "	100.
"ME"	3.13594
"n"	25.
" σ "	8.

- iii** $n = 50 \Rightarrow 95\% \text{ CI } [97.8, 102.2]$

zInterval 8,100,50,0.95: stat.results	
"Title"	"z Interval"
"CLower"	97.7826
"CUpper"	102.217
" $\bar{x}$ "	100.
"ME"	2.21745
"n"	50.
" σ "	8.

- iv** $n = 150 \Rightarrow 95\% \text{ CI } [98.7, 101.3]$

zInterval 8,100,150,0.95: stat.results	
"Title"	"z Interval"
"CLower"	98.7198
"CUpper"	101.28
" $\bar{x}$ "	100.
"ME"	1.28024
"n"	150.
" σ "	8.

- b** At the same significance level the larger sample size we take the narrower a confidence interval we get. In the formula $\left[\bar{x} - \frac{\sigma}{\sqrt{n}} \times z_{\frac{\alpha}{2}}, \bar{x} + \frac{\sigma}{\sqrt{n}} \times z_{\frac{\alpha}{2}} \right]$ the length of the interval is symmetrical with respect to the mean of the sample by the same term that is obtained by division by $\sqrt{n}$ therefore the larger n we take the smaller number we obtain.

Exercise 3C

- 1 a** 90% CI [14.45, 15.55]

tInterval 15,1,2,15,0.9: stat.results	
"Title"	"t Interval"
"CLower"	14.4543
"CUpper"	15.5457
" $\bar{x}$ "	15.
"ME"	0.545722
"df"	14.
"SX := S _{n-1} X"	1.2
"n"	15.

- b** 99% CI [-25.81, -20.19]

tInterval -23,5,8,32,0.99: stat.results	
"Title"	"t Interval"
"CLower"	-25.8135
"CUpper"	-20.1865
" $\bar{x}$ "	-23.
"ME"	2.81348
"df"	31.
"SX := S _{n-1} X"	5.8
"n"	32.

- c** 95% CI [3430, 3526]

tInterval 3478,429,310,0.95: stat.results	
"Title"	"t Interval"
"CLower"	3430.06
"CUpper"	3525.94
" $\bar{x}$ "	3478.
"ME"	47.9434
"df"	309.
"SX := S _{n-1} X"	429.
"n"	310.

- 2 a** 99% CI [1.94, 8.06]

tInterval {1,2,3,4,5,6,7,8,9},1,0.99: stat.results	
"Title"	"t Interval"
"CLower"	1.93696
"CUpper"	8.06304
" $\bar{x}$ "	5.
"ME"	3.06304
"df"	8.
"SX := S _{n-1} X"	2.73861
"n"	9.

- b** 95% CI [0.112, 0.159]

tInterval {0.1,0.12,0.15,0.18,0.13,0.12,0.09},1,0.95: stat.results	
"Title"	"t Interval"
"CLower"	0.111728
"CUpper"	0.159181
" $\bar{x}$ "	0.135455
"ME"	0.023726
"df"	10.
"SX := S _{n-1} X"	0.035317
"n"	11.

c 90% CI [319.6, 324.9]

CH3 Exercise C	
Interval {321,325,330,324,325,326,317,318}	
"Title"	"t Interval"
"CLower"	319.612
"CUpper"	324.921
"x"	322.267
"ME"	2.65433
"df"	14.
"sx := S _{n-1} x"	5.83667
"n"	15.

3 99% CI for the weight of mandarins is [67.6 g, 82.5 g]

CH3 Exercise C	
Interval {66,70,75,84,90,77,71,68,80,85,72}	
"Title"	"t Interval"
"CLower"	67.6363
"CUpper"	82.5304
"x"	75.0833
"ME"	7.44705
"df"	11.
"sx := S _{n-1} x"	8.30617
"n"	12.

4 a $\bar{x} = \frac{12.345 + 14.355}{2} = 13.35$

b $s = 1.5 \Rightarrow t_c \times \frac{1.5}{\sqrt{8}} = 13.35 - 12.345$

$\Rightarrow t_c = 1.895$

$\Rightarrow P(-1.895 < t < 1.895 | \nu = 7) = 0.900$

CH3 Exercise C	
$(13.35 - 12.345) \cdot \sqrt{8}$	1.89505
1.5	
$tCdf(-1.895046173579, 1.895046173579, 7)$	0.900069

The confidence level is 90%.

5 a 95% CI [483.4, 588.6]

CH3 Exercise C	
Interval 536,95,15,0,95: stat.results	
"Title"	"t Interval"
"CLower"	483.391
"CUpper"	588.609
"x"	536.
"ME"	52.6092
"df"	14.
"sx := S _{n-1} x"	95.
"n"	15.

b 99% CI [463.0, 609.0]

CH3 Exercise C	
"Title"	"t Interval"
"CLower"	462.981
"CUpper"	609.019
"x"	536.
"ME"	73.0187
"df"	14.
"sx := S _{n-1} x"	95.
"n"	15.

c For the same set of data, a higher significance level will mean a wider confidence interval. The result is expected since the confidence interval is symmetrical about the mean value

$$\bar{x} \pm \frac{s}{\sqrt{n}} \times t_c$$

Exercise 3D

1 a

	A	B	C	D
			=a[]-b[]	
1	15	18	-3	
2	17	15	2	
3	23	23	0	
4	17	19	-2	
5	18	15	3	
C1	=-3			

CH3 Exercise D	
"Title"	"t Interval"
"CLower"	-2.18091
"CUpper"	1.98091
"x"	-0.1
"ME"	2.08091
"df"	9.
"sx := S _{n-1} x"	2.02485
"n"	10.

99% CI [-2.18, 1.98]

b

	A	B	C	D	E
			=a[]-b[]		=d[]-e[]
8	1	93	102	-9	
9	-2	102	108	-6	
10	-1	106	109	-3	
11		99	111	-12	
12		85	103	-18	
F13					

CH3 Exercise D	
"Title"	"t Interval"
"CLower"	-13.7287
"CUpper"	-7.43794
"x"	-10.5833
"ME"	3.14539
"df"	11.
"sx := S _{n-1} x"	6.06717
"n"	12.

90% CI [-13.73, -7.44]

c

	A	B	C	D	E
			=d[]-e[]		=g[]-h[]
4	-8	0.58	0.69	-0.11	
5	-14	0.82	0.78	0.04	
6	-7	0.77	0.65	0.12	
7	-13	0.9	0.8	0.1	
8	-9	1.02	0.95	0.07	
I8	=0.07				

CH3 Exercise D ▾

"Title"	"t Interval"
"CLower"	-0.060895
"CUpper"	0.085895
"x̄"	0.0125
"ME"	0.073395
"df"	7
"sx := Sn-1x"	0.08779
"n"	8

6/99

95% CI [-0.069, 0.859]

2 a

Blood sample	1	2	3	4	5	6	7	8	9	10	11	12
Bob	144	153	170	183	125	95	148	177	160	155	170	135
Rick	141	161	173	174	119	104	135	175	164	158	167	142
Bob-Rick	3	-8	-3	9	6	-9	13	2	-4	-3	3	-7

b 95% CI [-4.27, 4.60]

CH3 Exercise D ▾

	A	B	C	D
8	177	175	2	
9	160	164	-4	
10	155	158	-3	
11	170	167	3	
12	135	142	-7	

C12 = -7

CH3 Exercise D ▾

	A	B	C	D
1	3	Title	t Interval	
2	-8	CLower	-4.26715	
3	-3	CUpper	4.60048	
4	9	x̄	0.166667	
5	6	ME	4.43381	

E1 = "t Interval"

Exercise 3E**1 a** $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0, \alpha = 0.1$

CH3 Exercise E ▾

zTest 10,2,12,20,0: stat.results

"Title"	"z Test"
"Alternate Hyp"	" $\mu \neq \mu_0$ "
"z"	4.47214
"PVal"	0.000008
"x̄"	12.
"n"	20.
"σ"	2.

1/99

Since the p -value is $0.000008 < 0.1$ we reject the null hypothesis at the 10% significance level.

b $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0, \alpha = 0.05$

CH3 Exercise E ▾

zTest 2,0,3,1,9,25,0: stat.results

"Title"	"z Test"
"Alternate Hyp"	" $\mu \neq \mu_0$ "
"z"	-1.66667
"PVal"	0.095581
"x̄"	1.9
"n"	25.
"σ"	0.3

2/99

Since the p -value is $0.095581 > 0.05$ we have no sufficient evidence to reject the null hypothesis at the 5% significance level.

c $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0, \alpha = 0.01$

CH3 Exercise E ▾

zTest -235,12,8,-238,119,0: stat.results

"Title"	"z Test"
"Alternate Hyp"	" $\mu \neq \mu_0$ "
"z"	-2.55673
"PVal"	0.010566
"x̄"	-238.
"n"	119.
"σ"	12.8

3/99

Since the p -value is $0.010566 > 0.01$ we have no sufficient evidence to reject the null hypothesis at the 1% significance level.

2 a $H_0: \mu = \mu_0, H_1: \mu < \mu_0, \alpha = 0.1$

CH3 Exercise E ▾

zTest 10,4,9,20,-1: stat.results

"Title"	"z Test"
"Alternate Hyp"	" $\mu < \mu_0$ "
"z"	-1.11803
"PVal"	0.131776
"x̄"	9.
"n"	20.
"σ"	4.

4/99

Since the p -value is $0.131776 > 0.1$ we have no sufficient evidence to reject the null hypothesis at the 10% significance level.

b $H_0: \mu = \mu_0, H_1: \mu < \mu_0, \alpha = 0.01$

CH3 Exercise E ▾

zTest 21,4,0.75,21.2,50,-1: stat.results

"Title"	"z Test"
"Alternate Hyp"	" $\mu < \mu_0$ "
"z"	-1.88562
"PVal"	0.029673
"x̄"	21.2
"n"	50.
"σ"	0.75

5/99

Since the p -value is $0.029673 > 0.01$ we have no sufficient evidence to reject the null hypothesis at the 1% significance level.

c $H_0: \mu = \mu_0, H_1: \mu < \mu_0, \alpha = 0.01$

CH3 Exercise E	
zTest -235,12.8,-238,119,-1: stat.results	
"Title"	"z Test"
"Alternate Hyp"	" $\mu < \mu_0$ "
"z"	-2.55673
"PVal"	0.005283
" $\bar{x}$ "	-238.
"n"	119.
" σ "	12.8
9/99	

Since the p -value is $0.005283 < 0.01$ we reject the null hypothesis at the 1% significance level.

3 a $H_0: \mu = \mu_0, H_1: \mu > \mu_0, \alpha = 0.05$

CH3 Exercise E	
zTest 10,5,12,20,1: stat.results	
"Title"	"z Test"
"Alternate Hyp"	" $\mu > \mu_0$ "
"z"	1.78885
"PVal"	0.036819
" $\bar{x}$ "	12.
"n"	20.
" σ "	5.
7/99	

Since the p -value is $0.036819 < 0.05$ we reject the null hypothesis at the 5% significance level.

b $H_0: \mu = \mu_0, H_1: \mu > \mu_0, \alpha = 0.1$

CH3 Exercise E	
zTest 27,3,3,6,28.,40,1: stat.results	
"Title"	"z Test"
"Alternate Hyp"	" $\mu > \mu_0$ "
"z"	1.22977
"PVal"	0.109391
" $\bar{x}$ "	28.
"n"	40.
" σ "	3.6
8/99	

Since the p -value is $0.109391 > 0.1$ we have no sufficient evidence to reject the null hypothesis at the 10% significance level.

c $H_0: \mu = \mu_0, H_1: \mu > \mu_0, \alpha = 0.01$

CH3 Exercise E	
zTest -73,3,3,72,-71,6,92,1: stat.results	
"Title"	"z Test"
"Alternate Hyp"	" $\mu > \mu_0$ "
"z"	3.60977
"PVal"	0.000153
" $\bar{x}$ "	-71.6
"n"	92.
" σ "	3.72
9/99	

Since the p -value is $0.000153 < 0.01$ we reject the null hypothesis at the 1% significance level.

- 4 a H_0 : "The mean weight is 26 g." ($\mu = 26$)
 H_1 : "The mean weight is not 26 g." ($\mu \neq 26$)
 b Use the z-test.

CH3 Exercise E			
worms		=zTest(26,	
1	22	Title	z Test
2	25	Alternate...	$\mu \neq \mu_0$
3	31	z	4.42627
4	35	PVal	0.00001
5	28	$\bar{x}$	30.
C4		=9.5958692475566E-6	

Since the p -value is $0.00001 < 0.01$ we reject the null hypothesis at the 1% significance level and conclude that the harvested snails are not from the population.

- 5 a H_0 : "The mean level of fat in the drink is 1.4 g." ($\mu = 1.4$)
 H_1 : "The mean level of fat in the drink is more than 1.4 g." ($\mu > 1.4$)

b Use the z-test.

CH3 Exercise E			
fat		=zTest(1.4	
1	1.43	Title	z Test
2	1.52	Alternate Hyp	$\mu > \mu_0$
3	1.35	z	0.424264
4	1.38	PVal	0.335687
5	1.42	$\bar{x}$	1.445
C		=zTest(1.4,0.3,a[...],1,1): CopyVar	

Since the p -value is $0.335687 > 0.05$ we have no sufficient evidence to reject the null hypothesis at the 5% significance level and conclude that the company's claim is correct.

- 6 a H_0 : "The mean volume of juice in the bottle is 300 ml." ($\mu = 300$)
 H_1 : "The mean volume of juice in the bottle is less than 300 ml." ($\mu < 300$)

b Use the z-test.

CH3 Exercise E			
volume		=zTest(30	
1	295	Title	z Test
2	288	Alternate Hyp...	$\mu < \mu_0$
3	293	z	-2.60364
4	301	PVal	0.004612
5	302	$\bar{x}$	295.75
C		=zTest(300,7.3,a[...],1,-1): CopyVar	

Since the p -value is $0.004612 < 0.1$ we reject the null hypothesis at the 10% significance level and conclude that the bottles contain less volume than stated.

Exercise 3F

- 1 a $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0, \alpha = 0.05$

"Title"	"t Test"
"Alternate Hyp"	"μ ≠ μ0"
"t"	-0.486504
"PVal"	0.638237
"df"	9.
"x̄"	4.8
"SX := Sn-1X"	1.3
"n"	10.

Since the p -value $0.6382 > 0.05$ we have no sufficient evidence to reject the null hypothesis at the 5% significance level.

- b $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0, \alpha = 0.1$

"Title"	"t Test"
"Alternate Hyp"	"μ ≠ μ0"
"t"	-1.66667
"PVal"	0.10858
"df"	24.
"x̄"	1.9
"SX := Sn-1X"	0.3
"n"	25.

Since the p -value $0.10858 > 0.1$ we have no sufficient evidence to reject the null hypothesis at the 10% significance level.

- c $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0, \alpha = 0.01$

"Title"	"t Test"
"Alternate Hyp"	"μ ≠ μ0"
"t"	3.48125
"PVal"	0.013122
"df"	6.
"x̄"	-35.3
"SX := Sn-1X"	0.532
"n"	7.

Since the p -value $0.013122 > 0.01$ we have no sufficient evidence to reject the null hypothesis at the 1% significance level.

- 2 a $H_0: \mu = \mu_0, H_1: \mu < \mu_0, \alpha = 0.05$

"Title"	"t Test"
"Alternate Hyp"	"μ < μ0"
"t"	-1.99172
"PVal"	0.027947
"df"	29.
"x̄"	14.2
"SX := Sn-1X"	2.2
"n"	30.

Since the p -value is $0.027947 < 0.05$ we reject the null hypothesis at the 5% significance level.

- b $H_0: \mu = \mu_0, H_1: \mu < \mu_0, \alpha = 0.01$

"Title"	"t Test"
"Alternate Hyp"	"μ < μ0"
"t"	-2.99871
"PVal"	0.007494
"df"	9.
"x̄"	119.8
"SX := Sn-1X"	2.32
"n"	10.

Since the p -value is $0.007494 < 0.01$ we reject the null hypothesis at the 1% significance level.

- c $H_0: \mu = \mu_0, H_1: \mu < \mu_0, \alpha = 0.1$

"Title"	"t Test"
"Alternate Hyp"	"μ < μ0"
"t"	-0.816497
"PVal"	0.225675
"df"	5.
"x̄"	622.8
"SX := Sn-1X"	12.6
"n"	6.

Since the p -value $0.225675 > 0.1$ we have no sufficient evidence to reject the null hypothesis at the 10% significance level.

- 3 a $H_0: \mu = \mu_0, H_1: \mu > \mu_0, \alpha = 0.1$

"Title"	"t Test"
"Alternate Hyp"	"μ > μ0"
"t"	-0.667483
"PVal"	0.743755
"df"	19.
"x̄"	0.95
"SX := Sn-1X"	0.335
"n"	20.

Since the p -value $0.743755 > 0.1$ we have no sufficient evidence to reject the null hypothesis at the 10% significance level.

- b $H_0: \mu = \mu_0, H_1: \mu > \mu_0, \alpha = 0.05$

"Title"	"t Test"
"Alternate Hyp"	"μ > μ0"
"t"	3.53553
"PVal"	0.004763
"df"	7.
"x̄"	26.4
"SX := Sn-1X"	1.12
"n"	8.

Since the p -value $0.004763 < 0.05$ we reject the null hypothesis at the 5% significance level.

- c $H_0: \mu = \mu_0$, $H_1: \mu > \mu_0$, $\alpha = 0.01$

Field	Value
Title	"t Test"
Alternate Hyp	" $\mu > \mu_0$ "
t	1.25463
PVal	0.115078
df	14
$\bar{x}$	758.6
SX := $S_{n-1}X$	14.2
n	15

Since the p -value $0.115078 > 0.01$ we have no sufficient evidence to reject the null hypothesis at the 1% significance level.

- 4 H_0 : "The mean volume is 120 ml." ($\mu = 120$)
 H_1 : "The mean volume is not 120 ml." ($\mu \neq 120$)

Field	Value
Title	"t Test"
Alternate Hyp	" $\mu \neq \mu_0$ "
t	0.283654
PVal	0.784882
df	7

Since the p -value $0.283654 > 0.01$ we have no sufficient evidence to reject the null hypothesis at the 1% significance level and conclude that the factory advertised a correct volume of a particular ice-cream product.

- 5 H_0 : "The mean life expectancy is 30 000 hours." ($\mu = 30\,000$)
 H_1 : "The mean life expectancy is less than 30 000 hours." ($\mu < 30\,000$)

Field	Value
Title	"t Test"
Alternate Hyp	" $\mu < \mu_0$ "
t	-1.51876
PVal	0.094643
df	5

Since the p -value $0.094543 < 0.1$ we reject the null hypothesis at the 10% significance level and conclude that the manufacturer claims a longer life expectancy of the LED lamps.

Exercise 3G

- 1 H_0 : "There is no difference in finishing times." ($\mu_d = 0$)
 H_1 : "There is a difference in finishing times." ($\mu_d \neq 0$)

Use the t -test on the difference of times on those two different cubes.

	cube1	cube2	diff
2	35	38	-3
3	41	40	1
4	30	34	-4
5	28	30	-2
6	46	44	2

Field	Value
Title	"t Test"
Alternate Hyp	" $\mu \neq \mu_0$ "
t	0.164399
PVal	0.874063
df	7

Since the p -value $0.874063 > 0.05$ we have no sufficient evidence to reject the null hypothesis at the 5% significance level and conclude that there is no difference in finishing times on the two Rubik's Cubes.

- 2 H_0 : "There is no difference in the scores." ($\mu_d = 0$)
 H_1 : "There is a difference in the scores." ($\mu_d \neq 0$)

Use the t -test on the difference of scores on the two types of dart.

	old	new	dartdiff
1	85	90	-5
2	92	95	-3
3	100	98	2
4	97	99	-2
5	89	93	-4

Field	Value
Title	"t Test"
Alternate Hyp	" $\mu \neq \mu_0$ "
t	-2.13719
PVal	0.085622
df	5

Since the p -value $0.085622 < 0.1$ we have to reject the null hypothesis at the 10% significance level and conclude that players score a better result by using the new type of dart.

- 3 H_0 : "There is no difference in the weights." ($\mu_d = 0$)
 H_1 : "Students who join the programme drop some weight." ($\mu_d > 0$)

Use the t -test on the difference of weights before and after the programme.

	before	after	wdiff
1	55	52	3
2	82	84	-2
3	63	61	2
4	69	69	0
5	65	62	3

	wdiff	Title	t Test	PVal	df
1	3	3	t Test		
2	-2	Alternate...	$\mu > \mu_0$		
3	2	t	1.23335		
4	0	PVal	0.121576		
5	3	df	11		

Since the p -value is $0.121576 > 0.05$ we have no sufficient evidence to reject the null hypothesis at the 5% significance level and conclude that there is no difference in weight before and after the programme.

Exercise 3H

- 1 a $H_0 : X \sim N(5, 0.4^2)$
 $\Rightarrow \alpha = 1 - P(4.2 \leq X \leq 5.8) = 0.0455$
- b $H_1 : X \sim N(4.5, 0.4^2) \Rightarrow$
 $\beta = P(4.2 \leq X \leq 5.8) = 0.773$

1-normCdf(4.2, 5.8, 5, 0.4)	0.0455
normCdf(4.2, 5.8, 4.5, 0.4)	0.772796

- 2 a $H_0 : X \sim B\left(50, \frac{1}{2}\right) \Rightarrow E(X) = 50 \times \frac{1}{2} = 25$
 $\alpha = 1 - P(X \leq 35) = 0.00130$
- b $H_1 : X \sim B\left(50, \frac{4}{7}\right) \Rightarrow \beta = P(X \leq 35) = 0.978$

1-binomCdf(50, 1/2, 0.35)	0.001301
binomCdf(50, 4/7, 0.35)	0.977867

- 3 a $H_0 : X \sim \text{Po}(45) \Rightarrow \alpha = 1 - P(X \leq 52) = 0.133$
- b $H_1 : X \sim \text{Po}(40) \Rightarrow \beta = P(X \leq 52) = 0.972$

1-PoissonCdf(45, 0.52)	0.132809
poissonCdf(40, 0.52)	0.971943

Review exercise

- 1 Use z -interval since the standard deviation is known and the value is 114 g.

- a 95% CI [602, 702]

Title	z Interval
CLower	602.038
CUpper	701.962
x	652
ME	49.9618
n	20
sigma	114

- b 99% CI [586, 718]

Title	z Interval
CLower	586.339
CUpper	717.661
x	652
ME	65.6609
n	20
sigma	114

- c $95\% < 99\% \Rightarrow [602, 702] \subset [586, 718]$

We notice that a higher significance level means a wider confidence level.

- 2 a

Patient	A	B	C	D	E	F	G	H	I	J
Analyzer I	235	352	410	280	341	325	428	388	272	310
Analyzer II	237	343	416	272	336	329	413	396	265	315
Difference	-2	9	-6	8	5	-4	15	-8	7	-5

- b H_0 : "There is no difference in potassium levels."
 $(\mu_a = 0)$
 H_1 : "There is a difference in potassium levels."
 $(\mu_a \neq 0)$

	a1	a2	dl
1	235	237	-2
2	352	343	9
3	410	416	-6
4	280	272	8
5	341	336	5

CH3 Review...ise

	C1	D1	F1	F2
1	=a1-a2		=tTest(0,d	
2	-2 Title		t Test	
3	9 Alternate...		$\mu \neq \mu_0$	
4	-6 t		0.76657	
5	8 PVal		0.46297	
6	5 df		9.	

E =tTest(0,d1[,1,0): CopyVar Stat.,

Since the p -value $0.46297 > 0.01$ we have no sufficient evidence to reject the null hypothesis and we conclude that there is no difference in measurement of the two types of biochemical analyzers.

$$\begin{aligned} \mathbf{3 \ a} \quad \bar{x} &= \frac{\sum_{i=1}^{15} x_i}{15} = \frac{80}{15} = \frac{16}{3} = 5.33 \\ s^2 &= \frac{\sum_{i=1}^{15} x_i^2 - 15 \times \left(\frac{16}{3}\right)^2}{14} = \frac{488 - \frac{1280}{3}}{14} \\ &= \frac{184}{42} = \frac{92}{21} = 4.38 \end{aligned}$$

$$\mathbf{b} \quad s = \sqrt{s^2} = \sqrt{\frac{92}{21}} = 2.09$$

CH3 Review...ise

```
tInterval(16/3, sqrt(92/21), 15, 0.99: stat.results)
```

Title	t Interval
"CLower"	3.72456
"CUpper"	6.94211
"x"	5.33333
"ME"	1.60877
"df"	14.
"sx := Sn-1x"	2.09307

1/99

99% CI [3.72, 6.94]

- c** In 99% of the cases the mean value of a sample of 15 observations taken from the population will fall within the confidence interval [3.72, 6.94].

4 a $\bar{x} = \frac{47.2 + 55.2}{2} = 51.2$

$$\begin{aligned} \mathbf{b} \quad \sigma^2 &= 25 \Rightarrow \sigma = 5 \Rightarrow z_{\frac{\alpha}{2}} \times \frac{5}{\sqrt{6}} \\ &= 55.2 - 51.2 \Rightarrow z_{\frac{\alpha}{2}} = \frac{4\sqrt{6}}{4} \\ &= 1.960 \Rightarrow \frac{\alpha}{2} = \Phi(1.960) = 0.975 \end{aligned}$$

The confidence level of the interval is 95%.

5 a $s = 1000 \text{ m}$, $v = 120 \text{ km/h} = \frac{100}{3} \text{ m/s}$,
 $t = \frac{s}{v} \Rightarrow t = \frac{1000}{\frac{100}{3}} = 30 \text{ s}$

$$\mathbf{b} \quad \bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = 30.97$$

$$s^2 = \frac{\sum_{i=1}^{10} x_i^2 - 10 \times \left(\frac{3097}{100} \right)^2}{9} = \frac{2681}{9000} = 0.298$$

CH3 Review...ise

```
> { 31.2,30.8,30.4,30.8,31.3,32.1,30.3,31.4,3};
```

30.97

```
varSamp({ 31.2,30.8,30.4,30.8,31.3,32.1,30.3,31.4,3});
```

0.297889

2/99

- c** H_0 : “The average time is 30 s.” ($\mu = 30$)
 H_1 : “The average time is more than 30 s.”
($\mu > 30$)

We use the t -test since the standard deviation is unknown.

CH3 Review...ise

tTest 30,{31.2,30.8,30.4,30.8,31.3,32.1,30.7}

"Title"	"t Test"
"Alternate Hyp"	" $\mu > \mu_0$ "
"t"	5.62011
"PVal"	0.000163
"df"	9.
" $\bar{x}$ "	30.97
"SX := $S_{n-1}X$ "	0.545792
"n"	10.

1/3

Since the p -value $0.000163 < 0.05$ we reject the null hypothesis and can conclude that the average time is more than 30 seconds. Therefore, the company sets up the speedometers to show a higher speed.

6 a $\bar{x} = \frac{204 + 216}{2} = 210$

$$\begin{aligned} \mathbf{b} \quad X &\sim N(210, 144) \Rightarrow \sigma = 12 \Rightarrow 1.960 \times \frac{12}{\sqrt{n}} \\ &= 216 - 210 \Rightarrow \sqrt{n} = 3.92 \Rightarrow n = 15.37 \end{aligned}$$

The sample size should be 16.

1.3 1.4 1.5 CH3 Review...ise

$$\frac{204+326}{2}$$

265

$$\text{nSolve}\left(\frac{\text{invNorm}(0.975) \cdot 12}{\sqrt{n}} = 216 - 210, n\right)$$

15.3658

2/99

- 7 a For the speed values we used the midpoints of the intervals.

	A	B	C	D
1	5	9	Title	=OneVar(a
2	15	56	$\bar{x}$	One-Var... 23.4667
3	25	47	Σx	3520
4	35	25	Σx^2	99150
5	45	13	$sx := s_n - 1x$	10.5383

	A	B	C	D
1	5	9	Title	=OneVar(v
2	15	56	$\bar{x}$	One-Var... 23.4667
3	25	47	Σx	3520
4	35	25	Σx^2	99150
5	45	13	$sx := s_n - 1x$	10.5383

i $\bar{x} = 23.5$

ii $s = 10.5$

- b i = 95% CI [21.8, 25.2]

Interval v,f,0.95: stat.results	
"Title"	"t Interval"
"CLower"	21.7664
"CUpper"	25.1669
" $\bar{x}$ "	23.4667
"ME"	1.70026
"df"	149
"sx := $s_n - 1x$ "	10.5383
"n"	150

- ii 90% CI [22.0, 24.9]

Interval v,f,0.9: stat.results	
"Title"	"t Interval"
"CLower"	22.0425
"CUpper"	24.8908
" $\bar{x}$ "	23.4667
"ME"	1.42417
"df"	149
"sx := $s_n - 1x$ "	10.5383
"n"	150

- c We notice that $[22.0, 24.9] \subset [21.8, 25.2]$, so a 90% confidence interval is a subset of a 95% confidence interval.

- 8 $H_0: \mu = 10, H_1: \mu < 10$, using the mean of a sample of size 5.

a $H_0: \mu = 10 \Rightarrow \bar{X} \sim N\left(10, \frac{2}{5}\right)$

i 10% for $N(0, 1)$ is -1.282 so $\frac{\bar{x} - 10}{\sqrt{\frac{2}{5}}}$

$= -1.282 \Rightarrow \bar{x} = 10 - 1.282 \times \sqrt{\frac{2}{5}} = 9.19$

The critical region is the interval $]-\infty, 9.19[$

nSolve(normCdf(-9.E999,x,10, $\sqrt{\frac{2}{5}}$)=0.1,x)	9.18948
nSolve(normCdf(-9.E999,x,10, $\sqrt{\frac{2}{5}}$)=0.05,x)	8.9597

ii 5% for $N(0, 1)$ is -1.645 so $\frac{\bar{x} - 10}{\sqrt{\frac{2}{5}}}$

$= -1.645 \Rightarrow \bar{x} = 10 - 1.645 \times \sqrt{\frac{2}{5}} = 8.96$

The critical region is the interval $]-\infty, 8.96[$

b $H_1: \mu = 9.3 \Rightarrow \bar{X} \sim N\left(9.3, \frac{2}{5}\right)$

i $\beta = P(\bar{X} > 9.19) = 0.569$

ii $\beta = P(\bar{X} > 8.96) = 0.705$

normCdf(9.1894753867941,9.E999,9.3, $\sqrt{\frac{2}{5}}$)	0.569364
normCdf(8.9597033202516,9.E999,9.3, $\sqrt{\frac{2}{5}}$)	0.704731

- c When the probability of a Type I error decreases from 10% to 5%, the probability of a Type II error increase from 0.569 to 0.705.

4

Statistical modeling

Skills Check

- 1 a $2E(Z) - 3E(Y) + 2E(X)$
 b $4\text{Var}(Z) - 9\text{Var}(Y) + 4\text{Var}(X)$
 c $E(X)E(Y)E(Z)$
- 2 a $x = 43.5, y = -31.75$,
 $\text{Var}(X) = 46.25, \text{Var}(Y) = 98.69$

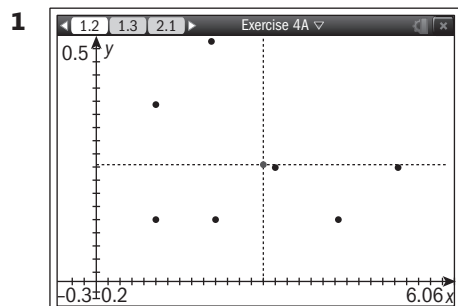
	B	C	D	E	F	G	H
=				=TwoVar'			
1	3	-21	Title	Two-va...			
2	9	-44	$\bar{x}$	43.5			
3	0	-23	Σx	174.			
4	2	-39	Σx^2	7754.			
5			$s_x := s_n...$	7.85281			
6			$\sigma_x := \sigma_n...$	6.80074			
7			n	4.			
8			$\bar{y}$	-31.75			
9			Σy	-127.			
10			Σy^2	4427.			
11			$S_y := S_n...$	11.471			
12			$\sigma_y := \sigma_n...$	9.93416			
13			Σxy	-5637.			

E1 = "Two-Variable Statistics"

- 3 a 0.382 b 0.951
 c 0.938 d 0.732

tCdf(-0.4,0.8,2)	0.382266
tCdf(-∞,1.83,10)	0.951412
tCdf(-1.75,∞,7)	0.938203
tCdf(-1.14,1.14,20)	0.732245

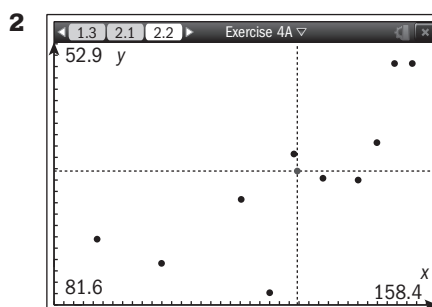
Exercise 4A



	A	B	C	D	E	F	G	H
=	xcoord	ycoord			=TwoVar'			=LinRegM
1	4	1	Title	Two-va...	Title			=Linear R...
2	1	1	$\bar{x}$	2.8	RegEqn	$m*x+b$		
3	3	2	Σx	28.	m	-0.3409...		
4	2	4	Σx^2	96.	b	2.95455		
5	1	3	$s_x := s_n...$	1.39841	r^2	0.146104		
6	2	4	$\sigma_x := \sigma_n...$	1.32665	r	-0.3822...		
7	5	2	n	10.	Resid	-0.5909...		
8	4	1	$\bar{y}$	2.				
9	4	1	Σy	20.				
10	2	1	Σy^2	54.				
11			$S_y := S_n...$	1.24722				
12			$\sigma_y := \sigma_n...$	1.18322				
13			Σxy	50.				

H1 = "Linear Regression (mx+b)"

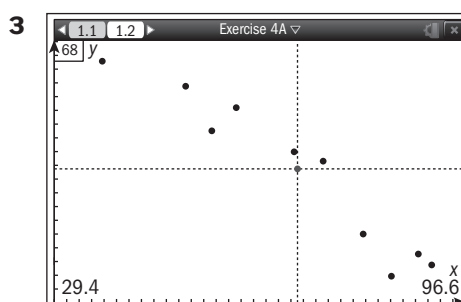
Since $r = -0.382$, there is a weak negative correlation between the two random variables.



	A	B	C	D	E	F	G	H
=	xcoord	ycoord			=TwoVar'			=LinRegM
1	100	33	Title	Two-Va...	Title	Linear R...		
2	115	38	$\bar{x}$	126.9	RegEqn	$m*x+b$		
3	140	40	Σx	1269.	m	0.245495		
4	149	51	Σx^2	165059.	b	9.44674		
5	88	36	$s_x := s_n...$	21.1421	r^2	0.630724		
6	132	40	$\sigma_x := \sigma_n...$	20.0572	r	0.794182		
7	152	51	n	10.	Resid	-0.9961...		
8	144	43	$\bar{y}$	40.6				
9	121	32	Σy	406.				
10	128	42	Σy^2	16868				
11			$S_y := S_n...$	6.53537				
12			$\sigma_y := \sigma_n...$	6.2				
13			Σxy	52509.				

B11

Since $r = 0.794$, there is a strong positive correlation between the two random variables.



	A	xcoord	B	ycoord	C	D	E	F	G	H
=							=TwoVar('			=LinRegM
1		55		58		Title	Two-Va...		Title	Linear R...
2		35		65		$\bar{x}$	66.7		RegEqn	$m \cdot x + b$
3		66		52		Σx	667.		m	-0.5559...
4		82		35		Σx^2	47645.		b	86.1833
5		91		36		$s_x := \sigma n \dots$	18.7264		r^2	0.939034
6		79		42		$\sigma_x := \sigma n \dots$	17.7654		r	-0.9690...
7		48		60		n	10.		Resid	2.39513.
8		52		55		$\bar{y}$	49.1			
9		71		50		Σy	491.			
10		88		38		Σy^2	25147.			
11						$S_y := \sigma n \dots$	10.744			
12						$\sigma_y := \sigma n \dots$	10.1926			
						Σxy	30995.			

Since $r = -0.970$, there is a strong negative correlation between the two random variables.

Exercise 4B

- $$\begin{aligned} \text{Cov}(X, Y) &= E(XY) - \mu_x \mu_y \\ &= E(XY) - \mu_y \mu_x \\ &= \text{Cov}(Y, X) \end{aligned}$$
- $$\begin{aligned} \text{Cov}(X, X) &= E(XX) - \mu_x \mu_x \\ &= E(X^2) - \mu_x^2 \\ &= \text{Var}(X) \end{aligned}$$
- $$\begin{aligned} \text{Cov}(aX, Y) &= E(aXY) - E(aX)E(Y) \\ &= aE(XY) - aE(X)E(Y) \\ &= a[E(XY) - (\mu_x \mu_y)] \\ &= a\text{Cov}(X, Y) \end{aligned}$$
- $$\begin{aligned} \text{Cov}(X, bY) &= E(X(bY)) - E(X)E(bY) \\ &= bE(XY) - E(X)E(bY) \\ &= bE(XY) - bE(X)E(Y) \\ &= b[E(XY) - (\mu_x \mu_y)] \\ &= b\text{Cov}(X, Y) \end{aligned}$$
- $$\begin{aligned} \text{Cov}(X_1 + X_2, Y) &= E[(X_1 + X_2)Y] - E(X_1 + X_2)E(Y) \\ &= E[X_1Y + X_2Y] \\ &\quad - [E(X_1) + E(X_2)]E(Y) \\ &= [E(X_1Y) - E(X_1)E(Y)] \\ &\quad + [E(X_2Y) - E(X_2)E(Y)] \\ &= \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y) \end{aligned}$$
- $$\begin{aligned} \text{Cov}(X, Y_1 + Y_2) &= E[(X)(Y_1 + Y_2)] - E(X)E(Y_1 + Y_2) \\ &= E[XY_1 + XY_2] - E(X)[E(Y_1) \\ &\quad + E(Y_2)] \\ &= [E(XY_1) - E(X)E(Y_1)] + [E(XY_2) \\ &\quad - E(X)E(Y_2)] \\ &= \text{Cov}(X, Y_1) + \text{Cov}(X, Y_2) \end{aligned}$$

- $$\begin{aligned} \text{Var}(X + Y) &= E(X + Y)^2 - [E(X + Y)]^2 \\ &= E(X^2 + 2XY + Y^2) - (\mu_x + \mu_y)^2 \\ &= E(X^2 + 2XY + Y^2) - \mu_x^2 - 2\mu_x\mu_y - \mu_y^2 \\ &= (E(X^2) - \mu_x^2) + (E(Y^2) - \mu_y^2) \\ &\quad + 2(E(XY) - \mu_x\mu_y) \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \end{aligned}$$
- Since X and Y are independent variables,

$$E(XY) = E(X)E(Y) = \mu_x \mu_y.$$

Hence, $\text{Cov}(X, Y) = E(XY) - \mu_x \mu_y = \mu_x \mu_y - \mu_x \mu_y = 0$.

Hence, $p = 0$.

Exercise 4C

	A	xcoord	B	ycoord	C	D	E	F	G
=							=LinRegT		
1		42		34	Title	Linear R...			
2		2		18	Alternate...	β & $p \neq \dots$			
3		26		14	RegEqn	$a + b \cdot x$			
4		22		20	t	4.05922			
5		15		41	Pval	0.002289			
6		44		44	df	10.			
7		50		51	a	9.76713			
8		38		45	b	0.73773			
9		12		17	s	9.69498			
10		28		27	SESlope...	0.181742			
11		22		28	r^2	0.622318			
12		1		1	r	0.788871			
13					Resid	{-6.7517...			

$$H_0: p = 0; H_1: p \neq 0$$

Since $0.00229 < 0.01$, we reject the null hypothesis. There is evidence of significant correlation between the two variables at the 1% level.

	A	xcoord	B	ycoord	C	D	E	F	G
=							=LinRegT		
1		64		69	Title	Linear R...			
2		66		64	Alternate...	β & $p \neq \dots$			
3		68		67	RegEqn	$a + b \cdot x$			
4		69		64	t	0.931872			
5		70		62	Pval	0.394177			
6		72		73	df	5.			
7		74		71	a	35.6			
8					b	0.457143			
9					s	4.10435			
10					SESlope...	0.490564			
11					r^2	0.147977			
12					r	0.384678			
13					Resid	{4.14285...			

$$H_0: p = 0; H_1: p = 0$$

Since $0.394 > 0.05$, there is not enough evidence to reject the null hypothesis. There is not evidence of significant correlation between the two variables at the 5% level.

3

	A	B	C	D	E	F	G
=					~LinRegT		
1	67	104		Title	Linear R...		
2	70	116		Alternate...	β & $p \neq \dots$		
3	70	121		RegEqn	$a+b*x$		
4	70	126		t	2.62304		
5	71	117		Pval	0.034257		
6	72	113		df	9.		
7	72	124		a	-93.		
8	72	126		b	3.		
9	73	127		s	5.78174		
10				SESlope...	1.14371		
11				r^2	0.49569		
12				r	0.704052		
13				Resid	[-4., -0.9...		

E1 ~"Linear Reg t Test"

$$H_0: p = 0; H_1: p = 0$$

Since $0.0343 < 0.1$, we reject the null hypothesis. There is evidence of significant correlation between the two variables at the 10% level.

Exercise 4D

1 a

	A	B	C	D	E	F	G	H	I	J
=						~LinRegM		~LinReg		~LinRegM
1	62	10	145	44.9	Title...	Linear R...	Title...	Linear R...	Title...	Linear R...
2	94	11	150	42.1	Reg...	$m*x+b$	Reg...	$m*x+b$	Reg...	$m*x+b$
3	81	8	139	30.2	m	0.014779	m	0.1642...	m	-0.2780...
4	62	9	148	41.4	b	8.48321	b	131.601	b	61.3778
5	58	8	130	41.8	r^2	0.035188	r^2	0.0426...	r^2	0.298465
6	86	10	150	38.4	r	0.187584	r	0.2064	r	-0.54632
7	59	11	154	54.1	Res...	{0.60049...	Res...	{3.216...	Res...	{0.75945...
8	72	9	140	38.6						
9	54	10	153	52.4						
10	60	9	120	38.6						
11										
12										
13										

E10

Column F shows the regression data for score versus age.

Column H shows the regression data for score versus height.

Column J shows the regression data for score versus weight.

$|r|$ is largest for the random variables score and weight, hence we use W .

b From the GDC, $S = -0.270W + 61.4$

c From the GDC,

	A	B	C	D	E	F	G
=							
1	Title...	Linear R...					
2	Alternate...	β & $p \neq \dots$					
3	RegEqn	$a+b*x$					
4	t	-1.84488					
5	PVal	0.102277					
6	df	8.					
7	a	61.3778					
8	b	-0.2780...					
9	s	6.1981					
10	SESlope...	0.150699					
11	r^2	0.298465					
12	r	-0.54632					
13	Resid	{0.75945...					

C13

Since $p = 0.102 > 0.05$ there is insufficient evidence to reject the null hypothesis, i.e. there is not evidence of significant correlation between the score and the weight at the 5% level.

2

	A	B	C	D	E	F	G
=					~LinRegM		
1	30	136		Title	Linear R...		
2	37	156		RegEqn	$m*x+b$		
3	38	150		m	1.90885		
4	32	140		b	82.763		
5	36	155		r^2	0.57174		
6	32	157		r	0.756135		
7	33	143		Resid	[-4.0286...		
8	38	160					
9	44	170					
10	38	144					
11							
12							
13							

F1

$r = 0.756$, which indicates a strong correlation between the two variables. Hence we find the line of regression, $y = 1.91x + 82.8$, and when $x = 35$ g, $y = 150$ mg.

3 a $H_0: p = 0; H_1: p = 0$

b

	A	B	C	D	E	F	G
=					~LinRegM		~LinRegT
1	120	80		Title	Linear R...	Title	Linear R...
2	125	90		RegEqn	$m*x+b$	Alternate...	β & $p \neq \dots$
3	130	92		m	0.62303	RegEqn	$a+b*x$
4	135	98		b	10.8182	t	10.0189
5	140	100		r^2	0.926185	Pval	0.000008
6	145	103		r	0.962385	df	8.
7	150	105		Resid	[-5.5818...	a	10.8182
8	155	108				b	0.62303
9	160	110				s	2.82414
10	165	110				SESlope...	0.062185
11						r^2	0.926185
12						r	0.962385
13						Resid	[-5.5818...

G1

$$r = 0.962; p = 8 \times 10^{-6}$$

c There is very strong evidence to indicate a positive association between the random variables X and Y .

d Since r is close to 1, we can find the regression line, $y = 0.623x + 10.8$, and when $x = 138$ mm, $y = 97$ mm.

$$4 \quad m = \frac{\sum xy - n\bar{x}\bar{y}}{\sum y^2 - n\bar{y}^2} = \frac{8390 - 5\left(\frac{182}{5}\right)\left(\frac{200}{5}\right)}{9850 - 5\left(\frac{200}{5}\right)^2} = 0.6$$

$$x - \bar{x} = 0.6(y - \bar{y}) \Rightarrow x = 0.6y + 12.4; x = 66.4 \text{ g}$$

5 For the gradient of the Y on X line of regression,

$$m = \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2} = \frac{580 - 10\left(\frac{30}{10}\right)\left(\frac{86}{10}\right)}{220 - 10\left(\frac{30}{10}\right)^2} = 2.48, \text{ hence}$$

$$y - \bar{y} = 2.48(x - \bar{x}) \Rightarrow y = 2.48x + 1.17$$

For the gradient of the X on Y line of regression,

$$m = \frac{\sum xy - n\bar{x}\bar{y}}{\sum y^2 - n\bar{y}^2} = \frac{580 - 10\left(\frac{30}{10}\right)\left(\frac{86}{10}\right)}{1588 - 10\left(\frac{86}{10}\right)^2} = 0.380, \text{ hence}$$

$$x - \bar{x} = 0.380(y - \bar{y}) \Rightarrow x = 0.380y - 29.6$$

Review exercise

- 1 a r is the unbiased estimate of ρ , and

$$r = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)(\sum y_i^2 - n\bar{y}^2)}} = 0.975$$

$$t = r\sqrt{\frac{n-2}{1-r^2}} = 10.7$$

Hence, since $p = .000039 < 0.05$, there is evidence of a strong positive relationship between the two random variables.

$$\text{b } m = \frac{\sum xy - n\bar{x}\bar{y}}{\sum y^2 - n\bar{y}^2} = \frac{37000 - 8\left(\frac{440}{8} \times \frac{606}{8}\right)}{49278 - 8\left(\frac{606}{8}\right)^2} = 1.088$$

$$x - \bar{x} = 1.088(y - \bar{y}) \Rightarrow x = 1.088y - 27.4;$$

$$x = 32.4$$

- 2 Since $\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$, the maximum

$\text{Cov}(X, Y)$ will occur when $\rho = 1$.

$$\text{Hence, } \text{Cov}_{\max}(X, Y) = \sqrt{17 \times 15} = 10.2$$

- 3 If $p = 0.9$ then $t = \text{inv}t(0.9, 18) = 1.33039$, and

$$\text{solving } 1.33039 = r\sqrt{\frac{18}{1-r^2}}, \text{ gives } r = 0.299$$

$$\begin{aligned} \text{4 a } \rho &= \frac{\text{Cov}(U, V)}{\sqrt{\text{Var}(U)\text{Var}(V)}} = \frac{\text{Cov}((X+Y), (X+Z))}{\sqrt{\text{Var}(X+Y)\text{Var}(X+Z)}} \\ &= \frac{\text{Cov}(XX) + \text{Cov}(XZ) + \text{Cov}(YX) + \text{Cov}(YZ)}{\sqrt{[\text{Var}(X+Y)][\text{Var}(X+Z)]}} \\ &= \frac{\text{Var}(X) + 0}{\sqrt{2\sigma^2}\sqrt{2\sigma^2}} = \frac{\sigma^2}{2\sigma^2} = \frac{1}{2} \end{aligned}$$

Hence it cannot be said that U and V are independent.

$$\begin{aligned} \text{b } \rho &= \frac{\text{Cov}(V, W)}{\sqrt{\text{Var}(V)\text{Var}(W)}} = \frac{\text{Cov}((X+Y), (X-Y))}{\sqrt{\text{Var}(X+Y)\text{Var}(X-Y)}} \\ &= \frac{\text{Cov}(XX) - \text{Cov}(XY) + \text{Cov}(YX) - \text{Cov}(YY)}{\sqrt{[\text{Var}(X+Y)][\text{Var}(X-Y)]}} \\ &= \frac{\text{Var}(X) - \text{Var}(Y)}{\sqrt{[\text{Var}(X+Y)][\text{Var}(X-Y)]}} = 0 \end{aligned}$$

We know that if V and W are independent then $\rho = 0$, but the converse does not necessarily hold.

5 a

	A	B	C	D	E	F	G
=				=LinRegT			
1	44.5	41.2	Title	Linear R...			
2	10.3	11.1	Alternate...	β & $p \neq \dots$			
3	20.1	18.7	RegEqn	$a+b*x$			
4	55	52.3	t	20.9954			
5	39.6	41.2	PVal	2.78035...			
6	24.1	26.5	df	8.			
7	31.2	29.3	a	2.59473			
8	9.5	11.2	b	0.904534			
9	22.3	25.1	s	1.9017			
10	35.1	33.2	SEslope...	0.043083			
11			r^2	0.982175			
12			r	0.991047			
13			Resid	{-1.6465...			
D5	=2.7803542327572E-8						

From the GDC, $r = 0.991$; $p = 2.78 \times 10^{-8}$

- b The p -value suggests a strong relationship between the two random variables, hence it makes sense to find the equation of the regression line: $y = 0.905x + 2.59$
- c $y = 0.905(19.8) + 2.59 = 20.5$
- d $p = 0.00620 < 0.01$, hence there is a significant correlation between the random variables.
- 6 a The gradient of the Y on X line, $y = a + bx$, is $\frac{\text{Cov}(X, Y)}{\text{Var}(X)}$, and the gradient of the X on Y line, $x = c + dy$, is $\frac{\text{Cov}(X, Y)}{\text{Var}(Y)}$. Since $r = 0$, $\text{Cov}(X, Y) = 0$, hence $y = a$ and $x = b$, and the two lines are perpendicular.
- b For $y = a + bx$, $m = b$, and for $x = c + dy$, $m = \frac{1}{d}$. $r = \pm 1 \Rightarrow r^2 = 1$. Since $r^2 = bd \Rightarrow bd = 1$, $b = \frac{1}{d}$. Hence, the two lines have the same gradient. Since both regression lines must go through $(\bar{x}, \bar{y})$, the lines are identical.
- c For the regression lines $y = a + bx$ and $x = c + dy$, $b = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$ and $d = \frac{\text{Cov}(X, Y)}{\text{Var}(Y)}$. Hence $bd = \left[\frac{\text{Cov}(X, Y)}{\text{Var}(X)\text{Var}(Y)} \right]^2 = r^2$. Since b and d are either both positive or both negative, $r = +\sqrt{bd}$ if both are positive, and $r = -\sqrt{bd}$ if both are negative.
- d Since b and d are both positive, $r = +\sqrt{bd} = \sqrt{0.19 \times 0.77} = 0.382$